STIF

NASA-TMX-58124

E7.5 - 10009

THE ERTS-1 INVESTIGATION (ER-600)

EARTH OBSERVATIONS DIVISION

SCIENCE AND APPLICATIONS DIRECTORATE

VOLUME VII — ERTS-1 LAND-USE ANALYSIS
OF THE HOUSTON AREA TEST SITE
TYPE III REPORT FOR PERIOD JULY 1972 — JUNE 1973

National Aeronautics and Space Administration

LYNDON B. JOHNSON SPACE CENTER

Houston, Texas

June 1974

RECEIVED
NOV 18 1974
SIS/902.6

VOLUME VII — ERTS-1 LAND-USE ANALYSIS

OF THE HOUSTON AREA TEST SITE

Type III Report for Period July 1972 — June 1973

Submitted by: _____

R. Bryan Erb
Principal Investigator
Applications Office
Earth Observations Division

Approved by: _____

R. B. MacDonald, Chief
Earth Observations Division

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

LYNDON B. JOHNSON SPACE CENTER

HOUSTON, TEXAS

June 1974

PREFACE

This report is one of seven separate reports prepared
by six discipline-oriented analysis teams of the Earth
Observations Division at the Lyndon B. Johnson Space Center
(JSC), Houston, Texas.

The seven reports were prepared originally for Goddard
Space Flight Center (GSFC) in compliance with require-
ments for the Earth Resources Technology Satellite (ERTS-1)
Investigation (ER-600). The project was approved and
funded by the National Aeronautics and Space Administration
(NASA) Headquarters in July 1972.

This report (Volume VII) was accomplished by the Land-
Use Analysis Team. The following members of the team were
personnel of the Earth Observations Division and the support
contractor:

W. P. Bennett, Lockheed Electronics Company, Inc.

C. M. Chesnutwood, JSC

J. G. Garcia, JSC

H. V. Johnson, Lockheed Electronics Company, Inc.

M. A. Lundelius, Lockheed Electronics Company, Inc.

D. P. McGuigan, Lockheed Electronics Company, Inc.

S. H. Tunnell, Lockheed Electronics Company, Inc.

The total investigation is documented in the following
reports:

Volume	Title	NASA Number
	A COMPENDIUM OF ANALYSIS RESULTS OF THE UTILITY OF ERTS-1 DATA FOR LAND RESOURCES MANAGEMENT	SP-347 JSC-08455

iv

Volume	Title	NASA Number
I	ERTS-1 AGRICULTURAL ANALYSIS	TMX-58117 JSC-08456
II	ERTS-1 COASTAL/ESTUARINE ANALYSIS	TMX-58118 JSC-08457
III	ERTS-1 FOREST ANALYSIS	TMX-58119 JSC-08458
IV	ERTS-1 RANGE ANALYSIS	TMX-58120 JSC-08459
V	ERTS-1 URBAN LAND-USE ANALYSIS	TMX-58121 JSC-08460
VI	ERTS-1 SIGNATURE EXTENSION ANALYSIS	TMX-58122 JSC-08461
VII	ERTS-1 LAND-USE ANALYSIS OF THE HOUSTON AREA TEST SITE	TMX-58124 / JSC-08463

CONTENTS

vi

TABLES

viii

FIGURES

x

ERTS-1 LAND-USE ANALYSIS OF THE
HOUSTON AREA TEST SITE

1.0 SUMMARY

1.1 OBJECTIVES

The general objective of this investigation was to
evaluate how well data from the ERTS-1 multispectral scanner
(MSS) could be used to detect, identify, and delineate
land-use features within the Houston Area Test Site (HATS),
an 18-county area around Houston established previously as
a land-use test area. A more specific objective was to
determine whether the land-use classification scheme
proposed in the U.S. Geological Survey (USGS) Circular 671
could be used as the basis for delineating land use by
conventional image interpretation and computer-aided
classification of ERTS-1 data.

1.2 ANALYTICAL APPROACH

An analysis of the entire 41,000-km^2 (16,000 mi^2) HATS
area was not feasible with the available man-hours and
computer time allotted to this investigation. Consequently,
a 4,660-km^2 (1,800-mi^2) study area was selected to correspond
to the data on one computer-compatible tape. This represented
a generally north-south-oriented area of one-fourth of a
scene of ERTS imagery.

An attempt was made to delineate Level-I land-use
categories by conventional image interpretation techniques.
These categories included urban and built-up land, agricul-
tural land, rangeland, forest land, nonforested wetland,

water and barren land. Black-and-white images of the study
area from ERTS-1 MSS bands 5 and 7 (October 4, 1972, pass)
were enlarged to a scale of approximately 1:250,000, and
delineations of land use were recorded on transparent over-
lays of these enlargements. A comparative study was conducted
by using similar interpretative techniques to delineate land-
use categories on enlargements made from first-generation color
composites obtained directly from a Data Analysis Station (DAS)
film recorder at JSC. These enlargements were also at a scale
of approximately 1:250,000 and were simulated color-infrared
composites of bands 4, 5, and 7. The same techniques were
used in delineating some Level-II land-use categories

Two basic computer-aided classification techniques
(supervised and nonsupervised) were employed in classifying
the study area into land-use categories. The Iterative
Self-Organizing Clustering System (ISOCLS), a nonsupervised
clustering algorithm, was used to group every sixth picture
element (pixel) from every sixth scan line into clusters of
pixels having similar spectral characteristics. This
reduction in the number of data points (45,630 pixels) was
necessary because the capacity of the computer was not
sufficient to process the total number of data points
(1.3 million pixels) covering the entire study area. The
3-percent, systematically aligned sample of data points,
uniformly distributed over the entire study area, was
grouped into spectral clusters. Each cluster represented a
portion of the full range of spectral variations found in the
study area. The input parameters to the cluster program
could be adjusted to provide more or fewer clusters.
However, after considering the amount of detail needed for

the proposed land-use hierarchy and the estimated computer
time required, it was reasoned that input parameters pro-
viding 13 clusters would be a reasonable compromise. Gray
map printouts depicting the spatial distribution of the
pixels grouped into each cluster were generated on the
computer. Each cluster was identified and assigned to a
specific land-use category by correlating the cluster
delineations on the gray maps with existing land-use maps
and aerial photographs and by analyzing pertinent cluster
statistics which had been plotted on graphs. After
grouping the clusters into the desired land-use categories,
a color-coded cluster map in the form of a color trans-
parency was produced on the JSC DAS film recorder.

Once the clusters had been grouped satisfactorily into
the Level-I land-use categories, the means and covariance
matrix statistics from the cluster analysis were substituted
for training field statistics as inputs in the LARSYS-II
supervised classification approach. (LARSYS is a set of
classification programs developed at the Laboratory for
Applications of Remote Sensing, Purdue University). The
use of cluster statistics in lieu of training field statistics
eliminated some of the difficulties which would have been
encountered in selecting representative training fields
for such a large study area where intensive ground truth
or large-scale aerial photography was not readily available
for analysis. Because of the relative spectral complexity
of much of the study-area landscape, it was deemed desirable
to be able to classify every pixel (instead of every sixth
pixel) within the entire study area. To do so, however, it
was necessary to divide the entire area into north-south
linear strips, with the number of data prints in each strip

not to exceed the storage capacity of the computer memory
drum. Level-I land-use classification maps of each strip
were subsequently mosaicked to form a classification map
of the entire study area. Experience gained in delineating
Level-I land-use categories by both supervised and non-
supervised classification techniques indicated that a
potential existed for dividing the urban and built-up category
into some Level-II categories. Some urban features (vegetated
residential areas) have spectral characteristics similar to
some nonurban features, such as forest and agricultural areas.
Because of this, it was necessary to reassign the 13 original
clusters into three Level-II urban and built-up categories
(residential, commercial-industrial-transportation, and
open) when a Level-II classification was made of only the
urbanized portion of the study area.

The accuracy of the three classification approaches was
assessed by measuring the agreement between the classified
data and base reference data established for the accuracy
analysis. Five accuracy test sites, ranging in size from
21 km^2 (8 mi^2) to 104 km^2 (40 mi^2), were established in
the study area. Base reference data were established by
visually classifying land use in each accuracy test site
from high-altitude, infrared-Ektachrome photography acquired
on April 22, 1972. Each site was divided into 2.6-km^2
quadrats, and the percent occurrence of each class in each
quadrat was measured using a dot-sampling technique. The
same procedure was performed on each class for each classi-
fication product, except for the computer-aided classifica-
tion maps where pixels in each class were counted and
converted to percent occurrence. The percent agreement

(class-by-class comparison of accuracy) between classifica-
tion products and base reference data was calculated, based
on the percent occurrence.

1.3 RESULTS

1. A visual comparison of all the classification
results shows a strong correlation in the areal patterns
of land use among the three analysis approaches used in
the investigation. However, there is a significant differ-
ence in detail. Because of the relatively small scale
(1:250,000) of the manually interpreted imagery, many of
the smaller features were difficult to portray. The result
is a pattern of relatively homogeneous tracts of land-use
classes.

2. The computer-aided classification maps display a
finer texture in the land-use patterns. This finer precision
is a result of the ability of the computer to classify each
pixel (about 0.45 ha).

3. The image interpreter can compensate for his in-
ability to resolve fine details with the ability to resolve
spatial patterns and relationships in the land-use features.
This was particularly true in the urban areas where many
linear features (e.g., secondary roads) could be visually
distinguished by conventional image interpretation, even
though the width of the features was well below the spatial
resolution threshold of the scanner.

4. Relatively high classification accuracies for Level-I
land-use categories were achieved by conventional image inte.-
pretation and computer-aided classification techniques, with
the exception of the urban and built-up category when it was

derived from computer classification of the entire study area. When only the preselected urban area was classified in Level-II categories, considerably better computer classification accuracies were attained. This apparent discrepancy in accuracies was probably due to the spectral heterogeneity of the urban scene in which vegetated urban features were spectrally similar to the vegetated agriculture-rangeland features.

1.4 CONCLUSIONS

1. It was concluded from this investigation that general land-use categories, as suggested for Level-I and some Level-II categories in the USGS Circular 671, could be obtained over relatively large areas from ERTS-1 MSS data by conventional image interpretation and computer-aided classification techniques.

2. In the computer-aided processing, a small (3 percent) sample of the available digital data was sufficient to identify the general land-use categories throughout the entire study area. This indicates that even larger geographic areas could be similarly classified without exceeding nominal computer capacities.

3. Where greater classification accuracies or more detailed land-use categorizations of larger areas are desired, it may become necessary to define categories of land use by geographic region, perform sampling within each region, and classify the entire large area into the desired land-use categories using computer-aided techniques.

2.0 INTRODUCTION

Traditionally, remote sensing techniques have been utilized for many years in mapping and inventorying features on the Earth's surface from aircraft altitudes. Most of these studies have been oriented toward rather narrow, specialized interests and have been concerned with relatively small segments of landscapes. With the increased concern about the use of land and resources, it seemed only natural that attempts should be made to extend remote sensing technology to orbital altitudes from which observations could be recorded of extensive landscapes having regional, or even continental, proportions. Prior to the July 23, 1972, launch of the Earth Resources Technology Satellite (ERTS-1), limited success had been achieved in mapping general land use from conventional and multiband photography acquired somewhat sporadically during the Gemini and Apollo missions. However, a new era in remote sensing of the Earth's surface was inaugurated when the ERTS-1 satellite was launched. Operating in a circular, Sun synchronous, near polar orbit, and equipped with more sophisticated multispectral sensors, it provided the ability to observe data recordings of any given point on the Earth's surface every 18 days. Since July 1972, the MSS subsystem has been routinely sensing the land surface in four spectral bands (0.5 to 1.1 μm) from an orbit of approximately 500 nmi (900 km) altitude.

3.0 OBJECTIVES

The general objective of this investigation was to evaluate how well data from the ERTS-1 MSS could be used to detect, identify, and delineate land-use features within the HATS 18-county area. A more specific objective was to determine whether the land-use classification scheme proposed in the USGS Circular 671 could be used as the basis for delineating land use by conventional image interpretation and computer classification of the ERTS-1 data.

The following were considered limiting factors in developing the scope of this investigation:

1. Predicted resolution limitation of the MSS.

2. Expected computer capacity.

3. Man-hours allotted to the project.

4. The broad connotation of the term "land use" made it desirable to emphasize initially only the Level-1 land-use classification found in Circular 671 applied to a specified study area within HATS.

As experience was gained in processing the ERTS-1 MSS data, the scope was broadened to include the evaluation of Level-II land-use classifications for limited areas of interest within the specified study area.

4.0 STUDY AREA

It was recognized that it would not be advisable to
attempt a computerized land-use classification of the entire
HATS area, 41,000 km^2 (16,000 mi^2), shown in figure 4-1,
with the estimated man-hours and computer time allotted to
this project. Consequently, a 4,660-km^2 (1,800-mi^2) study
area was selected to correspond to the data on one computer-
compatible tape. This represented a north-south oriented
rectangular area along the orbital track of approximately
one-fourth of a normal scene of ERTS-1 imagery. The
dimensions of the study area are approximately 40×115 km
(25×72 statute mi), shown in figure 4-1.

4.1 STUDY AREA DESCRIPTION

The study area, because of its north-south linear
extent, crosses a variety of landscapes. In the northern
portion are found heavily forested, very gently sloping
interfluves. Along the southwestern edge is a portion of a
major, rapidly expanding metropolitan area. The central
portion of the study area contains considerable agricultural
activities based upon the level, grass-covered coastal plain
soils. Toward the southern end of the study area the land-
scape merges into coastal rangelands, beaches, marshes,
forested wetlands, and numerous bays and estuaries character-
istic of broad, low-lying coastlines.

Although only a portion of the Houston metropolitan
complex is included within the study area, there is still a
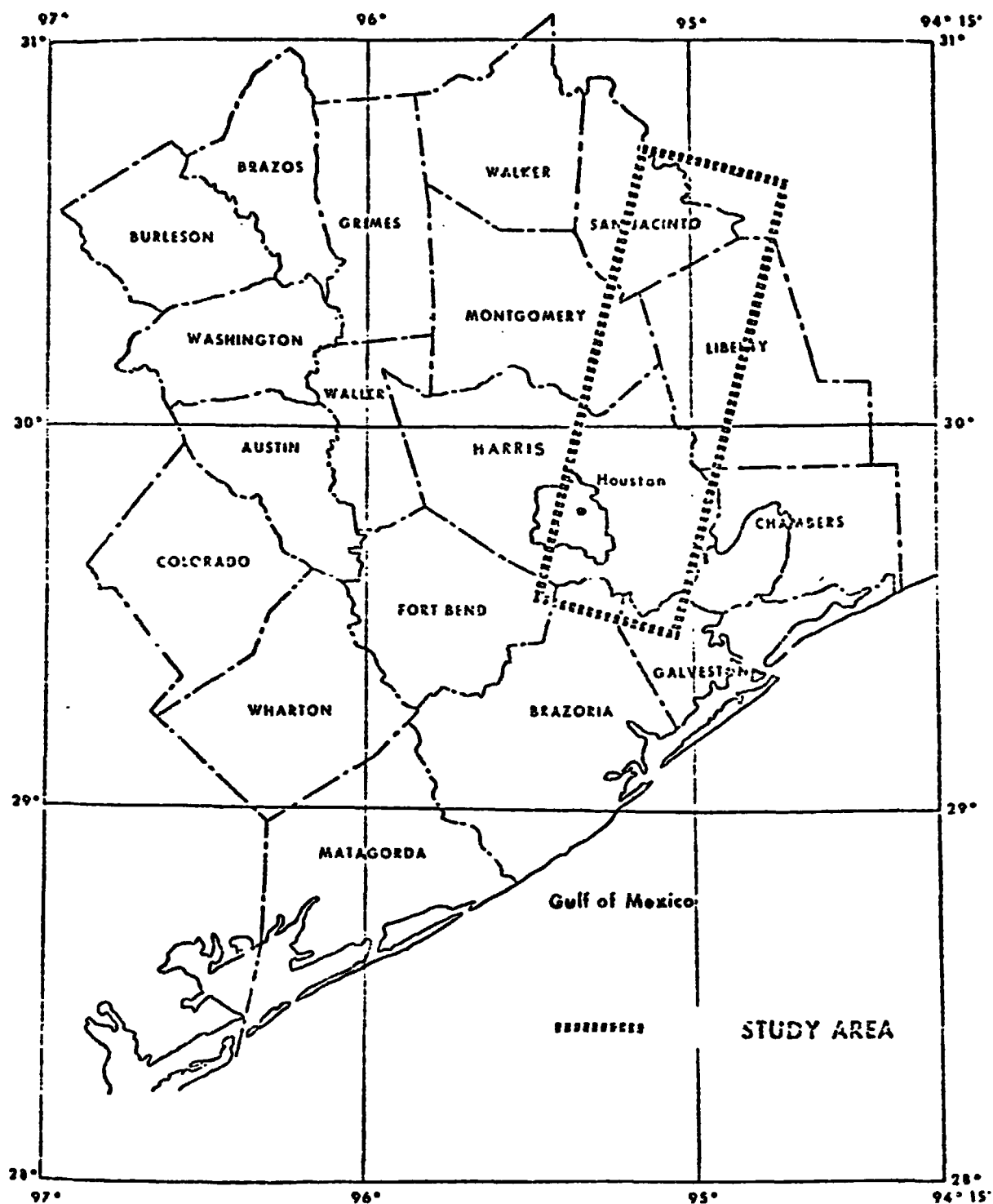
Figure 4-1.- HATS study area.

large variety of urban features. Some of these features
include:

　　1. A large, central business district.

　　2. Distinctive commercial strips along major
thoroughfares.

　　3. Large and small industrial complexes.

　　4. All types of residential areas ranging from low-
density, single-family dwellings to large apartment
complexes.

　　A variety of agricultural features is found in the
study area, including large homogeneous cultivated fields
and extensive pastures. Forest cover for the most part is
mixed deciduous interspersed with smaller stands of scattered
evergreen conifers. The major rivers, streams, and bayous
are lined with water-tolerant hardwoods; and tne adjacent
wet areas are flanked with mixed hardwoods and softwoods

4.2 LAND-USE HIERARCHY IN STUDY AREA

　　Although this investigation was based upon the objective
of using the land-use classification scheme from Circular 671,
it was recognized immediately that certain modifications
would be required to adapt it to the land-use features
found in the study area. The basic scheme from Circular 671
is shown in table IV-1. Of the nine Level-I categories in
the basic scheme, three categories (barren land, tundra, and
permanent snow and icefields) were not relevant because
these features do not exist in the study area, except for a
few very narrow strips of barren land in some of the stream
channels. It was expected that certain land-use features
in some of the other categories (for example, rangeland and
agricultural land) would have similar spectral characteristics

TABLE IV-1. — LAND-USE CLASSIFICATION SYSTEM FOR USE WITH REMOTE SENSOR DATA

Level I	Level II
Urban and Built-up Land	Residential
	^ommercial and Services
	Industrial
	Extractive
	Transportation, Communications, and Utilities
	Institutional
	Strip and Clustered Settlement
	Mixed
	Open and Other
Agricultural Land	Cropland and Pasture
	Orchards, Groves, Bush Fruits, Vineyards, and Horticultural Area
	Feeding Operations
	Other
Rangeland	Grass
	Savannas (Palmetto Prairies)
	Chaparral
	Desert Shrub
Forest Land	Deciduous
	Evergreen (Coniferous and Other)
	Mixed
Nonforested Wetland	Vegetated
	Bare
Water	Streams and Waterways
	Lakes
	Reservoirs
	Bays and Estuaries
	Other
Barren Land	Salt Flats
	Beaches
	Land Other than Beaches
	Bare Exposed Rock
	Other
Tundra	Tundra
Permanent Snow and Icefields	Permanent Snow and Icefields

These spectral similarities required modification of the
basic scheme if compatibility with ERTS-1 spectral data
was to be achieved. Further modification of the basic
scheme was also anticipated because it had been designed
primarily for remote sensors in general, rather than for
a specific type of sensor. The basic scheme was structured
for conventional interpretation utilizing both spatial
and spectral characteristics of land-use features, whereas
only spectral characteristics of features could be consid-
ered when automated data analysis procedures were applied
to the ERTS-1 data.

A land-use map with 20 categories had been constructed
of the HATS area prior to the publication of Circular 671 by
interpreting high altitude (60,000-ft or 13.3-km) color aerial
photography obtained in 1970. In order to use the HATS land-
use map as ground truth base for this investigation, it was
necessary to regroup some of the categories to be more
compatible with the Level-I categories (section 7) of
Circular 671. Some land-use definitional problems existed
between the two land-use schemes, so a few categories
(pasture versus rangeland, forest brushland versus range-
land) were not directly comparable. Consequently, it
was expected that problems would be encountered in deter-
mining the accuracies of the land-use classification
schemes that would be used in this investigation.

5.0 DATA UTILIZATION

The data used in this investigation included photographic materials, computer-compatible tapes, computer output materials, ground survey measurements, and other ancillary information.

The ERTS-1 imagery used in this project consisted of 70-mm black-and-white transparencies of the Houston area (frame 1037-16244, dated August 29, 1972, and frame 1073-16244, dated October 4, 1972) from each of the spectral bands of the MSS. Each spectral band frame was enlarged to a scale of approximately 1:1,000,000 in the form of black-and-white paper prints and film transparencies. Black-and-white paper prints and film transparency enlargements of approximately 1:250,000 scale were also used in this investigation. A limited number of color composites (paper and transparencies of 1:1,000,000 scale) was acquired from the GSFC later in the program.

One MSS computer-compatible tape dated August 29, 1972, containing data of one-fourth of an image frame was used as the basis for the investigation of computerized classification techniques in this project. This tape also was used for generating false-color composites on which conventional image interpretation techniques were supplied.

High- and low-altitude aircraft data over the selected study areas were used for classification verification.

6.0 ANALYTICAL APPROACH

The following two basic analytical approaches and one accuracy evaluation approach were used in the attempt to meet the objectives of this investigation:

1. Conventional image interpretation techniques were used as one approach in analyzing the ERTS-1 black-and-white imagery and false-color composite imagery generated from the digital data.

2. Both supervised and nonsupervised computerized classification procedures were used as another approach to analyze the ERTS-1 MSS digital data.

3. A statistical sampling approach was used to evaluate the accuracies that could be achieved by the analytical approaches in classifying the ERTS-1 MSS data into selected land-use categories.

6.1 CONVENTIONAL IMAGE INTERPRETATION APPROACH

Black-and-white imagery of bands 4, 5, 6, and 7 provided by GSFC was reviewed independently to establish which bands were best suited for the categories being studied within the land-use study area. Band 5 was selected for high reflectance categories such as highways, built-up areas, and some agricultural areas. Bank 7 best depicted water and hydrologic features. Transparencies from bands 5 and 7 were enlarged to approximately 1:250,000 scale, and land use was interpreted using conventional image interpretation techniques.

A translucent, stable base mylar overlay was keyed to
the enlarged band 7 transparency, and the Level-I category
water was delineated. The same overlay was then keyed to
the enlarged band 5 transparency, and the remaining Level-I
categories were delineated. Figure 6-1 shows an example of
band 5 imagery over the study area in which the land-use
delineations were made. Interpretation involved identifica-
tion of known signatures for each class and then extending
these signatures by interpreting image tone, texture, shape,
size, shadows, locations, and patterns.

Identical image interpretation techniques were used in
analyzing the false-color composite imagery generated from
the digital tape. However, two computer processing steps
had to be performed before the color composites were ready
for analysis. In the first step the ERTS-1 bulk data tape
was run through the EMBEDT program which converts the bulk
MSS tape to a tape format compatible with the DAS system.
The output of the EMBEDT program was a histogram of the
number of occurrences of each possible relative radiance
value of each of the four ERTS-1 channels. The histograms
were used to develop inputs to the second processing step.

The second processing step used the JSC program on the
DAS. This program produced a three-band JSC color composite
image of the ground scene for viewing and also for film
generation. The normal procedure was to compute the bias
and gain value for each band used as a function of the
histogram. The following equations were used:

$$\text{Bias} = \text{Min} \qquad \text{Gain} = \frac{255}{\text{Max-Min each band}}$$

Figure 6-1.— Base data for manual interpretation of land use.

BASE DATA FOR MANUAL INTERPRETATION OF LAND USE IN THE HOUSTON AREA

ERTS-1 MSS BULK DATA, BAND 4 (13.5X ENLARGEMENT) OCTOBER 4, 1972

6-3

The preceding computation resulted in a bias and gain value for
each band which produced the most detail available for the
scene as a whole. Note that this does not say that each
individual feature has the most detail available; it may or
may not. However, for the conglomerating of each individual
feature into a whole scene, it represents the greatest amount
of detail available.

After viewing the study area input and manipulating the
gain and bias controls, the following settings were selected:

	Band 4	Band 5	Band 7
Gain	7.536	6.600	7.310
Bias	-18	-10	0

These gain and bias settings were used to produce the three-
band JSC color composite film strip (fig. 6-2) used in
this analysis. The original film dimensions of the image
area were 7-1/4 inches by 26 inches with a nominal scale of
approximately 1:250,000.

Using conventional image interpretation procedures,
both Level-I and Level-II (urban built-up only) land-use
classifications were made from the JSC color composite film
transparency. Each category was annotated on a stable base
mylar overlay and keyed to the film strip of the prime study
area. The interpreter analyzed both spatial and spectral
patterns in defining specific categories. Reference was
made to the black-and-white 9- by 9-inch positive trans-
parencies when interpretation problems were encountered.

The results of the conventional image interpretation
approach are reported in section 7-1.
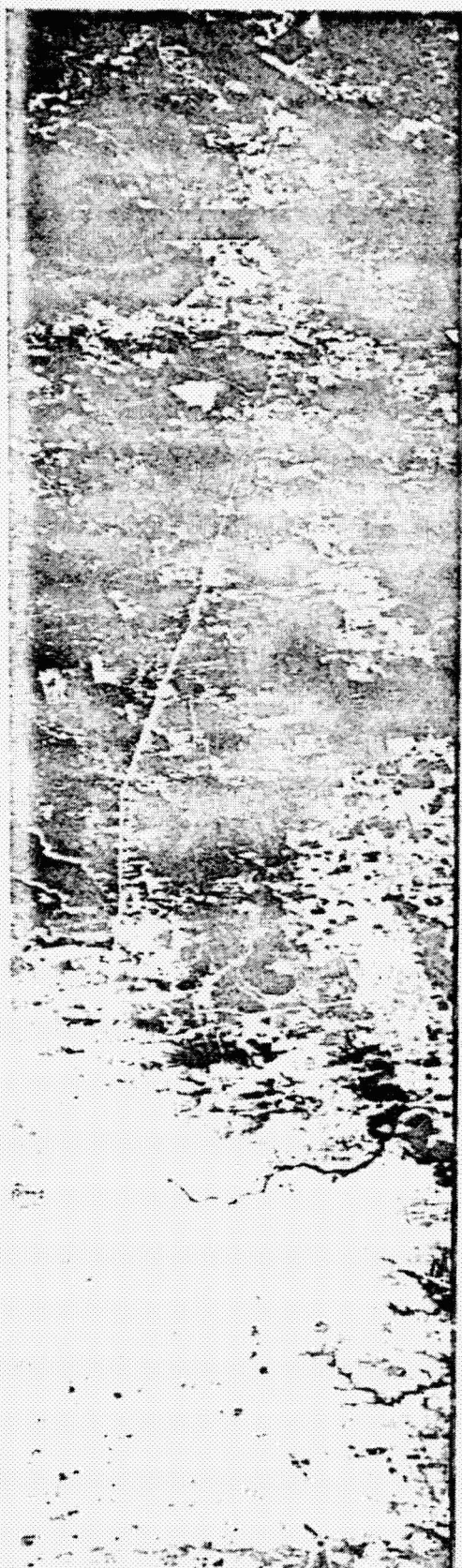
Figure 6-2.— JSC color composite of land-use study area.

6.2 COMPUTER CLASSIFICATION APPROACH

Nonsupervised clustering and supervised classification
algorithms were utilized to achieve a computer classifica-
tion of land use in the study area. The nonsupervised
clustering used the ISOCLS (Minter, 1972) clustering program
primarily to generate the "training" statistics (i.e., mean
and covariance matrices for subclasses of the land-use
hierarchy) required to perform the supervised classification.
ISOCLS statistics were analyzed to assign each cluster
generated by the program to specific land-use classes.

The LARSYS pattern recognition algorithm, developed by
the Laboratory for Applications of Remote Sensing, 1968, and
Ratcliff, 1970, was used for the supervised classification.
With training statistics for each class input from ISOCLS,
every pixel of ERTS MSS digital data over the study area
was assigned to a class based on the maximum likelihood ratio.
The resulting classification tape was then converted to a
film transparency which constitutes the classification map.

A sampling procedure was performed in order to obtain
a representative, workable sample of pixels from the study
area. The pixel sample was input into ISOCLS, and the
resulting clusters analyzed. The cluster statistics were
then submitted in lieu of training field statistics to the
LARSYS-II classification algorithm, and a land-use classifi-
cation of the study area was performed.

Each of the steps performed in the computerized approach
is discussed in detail in the following paragraphs. The
results of using the computerized approach are reported in
section 7.2.

6.2.1. Nonsupervised Clustering

The clustering algorithm, ISOCLS, is a "nonsupervised" iterative procedure which groups data of similar characteristics into distinct sets or clusters. The program requires certain input parameters which control the several group characteristics such as size, number of classes, and distance between groups before splitting or combining. The data, in this case, are the four spectral readings (one for each ERTS band) for each pixel. Based on the four-dimensional vector that describes each pixel, each pixel is assigned to a cluster: the mean is calculated for each cluster. A cluster is deleted if it has fewer than a specified number of points. The process of combining and/or reassigning data points continues until the desired number of iterations has been performed (Minter, 1972).

6.2.1.1 Sampling procedure.— Initially it was necessary to devise a sampling procedure to determine the spectral characteristics of the ground scene without having to consider all of the information contained within each and every pixel. Sampling was essential to reduce computer processing time because of the large amount of digital data contained on one magnetic tape from GSFC. The ground scene of one tape consists of approximately 1.9 million pixels, and each pixel is made up of a reading from each of the four ERTS-1 MSS bands or about 7.6 million readings.

The computer used for the ISOCLS program has a restriction on the amount of MSS data that it can accept. The number of MSS readings times the number of scanner channels cannot exceed 786,432. Obviously the field size of a

quarter-frame magnetic tape exceeds the capacity of the
computer program, and so the input requirement had to be
reduced accordingly. To achieve an adequate reduction in
scanner readings, a uniformly distributed sample of pixels
from the entire study area was used as input to the ISOCLS.
Every sixth pixel on every sixth scan line was designated
a sample point. This sample of 45,630 pixels represented
approximately 2.78 percent of the total pixels available in
the study area scene. The ISOCLS program already contained
the software implementation for selecting the uniformly
distributed sample of pixels.

6.2.1.2 Cluster procedure.— The next step entailed a limited
parametric study of the number of clusters output as a
function of the values assigned to the input parameter STDMAX
(i.e., the value of the standard deviation before a class is
split into two groups). The DLMIN (the minimum distance
threshold for combining clusters) input parameter was set
equal to 1.0, and the maximum number of iterations was set
equal to 10. Values of 0.8, 0.9, 1.0, and 1.1 were selected
as STDMAX values. Computer runs were set up and executed
using the modified ISOCLS program. The number of clusters
produced for each value of STDMAX follows:

STDMAX	Clusters
0.8	24
0.9	21
1.0	18
1.1	13

A study of the computer gray-map printouts revealed
that the grouping of sample pixels into 13 clusters appeared

to produce a map that was the best representation of actual land-use patterns in the study area. Generating more clusters gave the advantage of dividing the scene into more detailed spectral-based patterns but concurrently presented the disadvantages of requiring not only more computer processing time but also more time to aggregate the smaller clusters into larger, more meaningful patterns of land use. These observations led to the decision to use the STDMAX value of 1.1 (which resulted in 10 iterations and 13 clusters) for the subsequent computer clustering runs.

Tables and graphs were compiled from the statistics generated by processing the August 29, 1972, ERTS-1 data with the ISOCLS clustering algorithm and using the STDMAX value of 1.1 as input. These tables and graphs were studied as aids in aggregating the clusters into groups which had similar statistical characteristics and which were distributed in patterns resembling known land-use patterns in the study area. Table VI-1 shows the number of pixels within the sample that were assigned to each cluster. Tables VI-2 and VI-3 list the means and standard deviations of the counts of all the pixels assigned to each cluster. A count refers to the gray-scale value related to scene radiance from a resolution element within a spectral band. The radiance is measured in increments of 1, with a range of 0 to 127 in bands 4, 5, and 6, and 0 to 63 in band 7. (The higher the count, the greater the spectral radiance.)

The mean radiances (gray-scale mean values) as listed in table VI-2 were plotted on graphs to facilitate cluster interpretation (fig. 6-3). The four ERTS bands were plotted along the X-axis. The Y-axes are incremented in

TABLE VI-1. — DISTRIBUTION OF PIXELS (AUGUST 29, 1972, DATA)

Cluster	Pixels in Cluster	Percent
1	1 436	3.1
2	7 491	16.4
3	1 683	3.6
4	19 508	42.7
5	27	0.0
6	7 196	15.7
7	375	0.8
8	2 536	5.5
9	3 124	6.8
10	1 416	3.1
11	427	0.9
12	321	0.7
13	90	0.1

Total Number of
Pixels   =   45 630

TABLE VI-2. — MEAN RADIANCE FOR CLUSTERS
(AUGUST 29, 1972, DATA)

		Means		
Clusters	Band 4	Band 5	Band 6	Band 7
1	23.94	12.76	9.30	2.14
2	32.19	23.79	45.59	24.53
3	34.01	27.84	37.89	18.73
4	23.78	13.53	34.60	20.14
5	72.67	80.44	75.89	31.93
6	27.66	19.15	41.16	22.77
7	27.95	19.63	25.79	11.55
8	30.23	20.72	52.72	29.90
9	37.73	30.91	47.84	24.60
10	42.99	40.13	47.17	22.12
11	50.60	49.24	52.03	23.37
12	32.19	24.16	15.02	3.62
13	58.42	60.47	60.46	26.20

TABLE VI-3. — STANDARD DEVIATIONS OF CLUSTER RADIANCE
(AUGUST 29, 1972, DATA)

Standard Deviations

Cluster	Band 4	Band 5	Band 6	Band 7
1	1.15	1.41	1.54	1.39
2	2.58	2.12	2.14	1.56
3	4.28	3.74	3.36	2.38
4	2.11	1.54	2.71	1.70
5	6.66	6.62	4.99	2.64
6	2.65	2.31	2.57	1.81
7	2.71	3.15	4.01	3.09
8	2.81	2.89	4.88	3.62
9	2.43	2.37	3.60	2.35
10	2.85	2.60	4.93	3.47
11	3.40	3.88	4.93	3.48
12	2.46	3.32	2.91	1.22
13	4.65	4.21	4.31	2.87

gray-scale counts. The mean count of each cluster for every band was plotted at the midpoint of the range of the corresponding band. The curves appeared to fall into two basic groups or families — one representing vegetated surfaces and one representing nonvegetated surfaces. Further subdivision of the two basic families appeared feasible. The nonvegetated family was divided into urban (which would have to include bare land surfaces) and water. The vegetated family was divided into agriculture/rangeland, forest, and nonforested wetland (a hybrid group that appeared to be an integration of water and nonforested vegetative surfaces).

The weighted mean distances between each of the 13 clusters (table VI-4) and the standard deviations for each band in each cluster were inspected for possible indications that would demonstrate intercluster relationships. It was
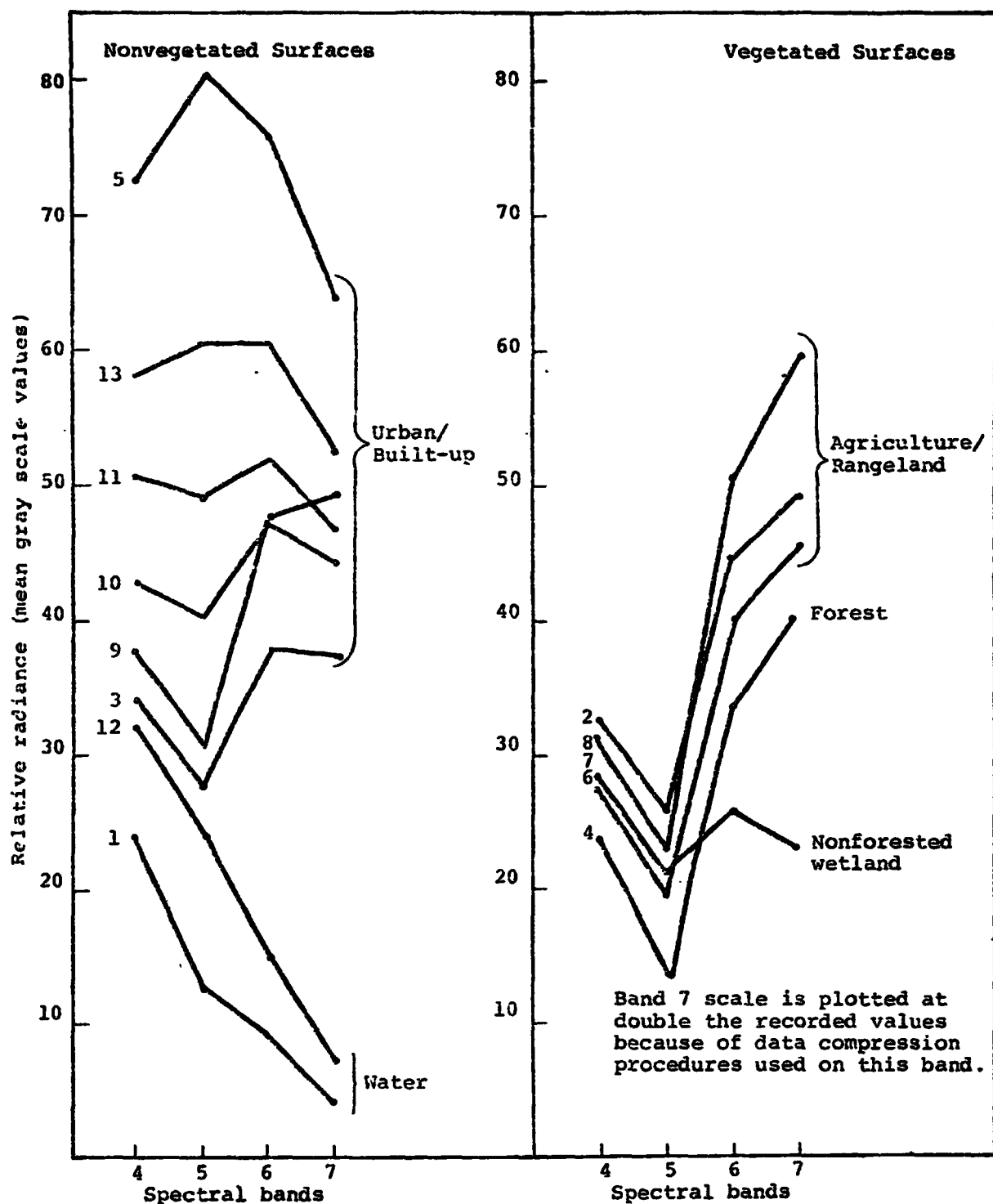
Figure 6-3.— Relative radiance values of clusters.

TABLE VI-4. — DISTANCE BETWEEN CLUSTER CENTERS

(AUGUST 29, 1972, DATA)

Cluster	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0.00	13.19	8.21	8.48	16.99	10.39	4.09	9.07	11.53	11.17	12.06	3.64	14.83
2	13.19	.00	2.17	4.34	9.14	1.74	4.43	1.64	1.98	4.04	5.38	9.92	7.21
3	8.21	2.17	.00	3.23	8.17	2.04	2.62	2.92	2.06	2.74	4.08	5.72	5.91
4	8.48	4.34	3.23	.00	12.02	2.25	2.74	3.67	5.93	7.72	8.65	7.56	10.48
5	16.99	9.14	8.17	12.02	.00	9.90	10.74	8.21	7.83	6.74	4.80	12.93	3.00
6	10.39	1.74	2.04	2.25	9.00	.00	3.28	2.11	3.43	5.18	6.38	8.21	8.12
7	4.09	4.43	2.62	2.74	10.74	3.28	.00	4.08	4.70	5.38	6.59	2.76	8.55
8	9.07	1.64	2.92	3.67	8.21	2.11	4.08	.00	2.63	4.37	5.43	7.44	6.83
9	11.53	1.98	2.06	5.93	7.83	3.43	4.70	2.63	.00	2.15	3.72	8.09	5.61
10	11.17	4.04	2.74	7.72	6.72	5.18	5.38	4.37	2.15	.00	1.94	7.11	4.02
11	12.06	5.38	4.08	8.65	4.80	6.38	6.59	5.43	3.72	1.94	.00	8.18	1.99
12	3.64	9.92	5.72	7.56	12.93	8.21	2.76	7.44	8.09	7.11	8.18	.00	10.53
13	14.83	7.21	5.91	10.48	3.00	8.12	8.55	6.83	5.61	4.02	1.99	10.53	.00

decided to group all clusters which had an interclass dis-
tance less than 4.0. Table VI-5 is the result of the grouping.
Each cluster was inspected and analyzed in terms of the
number of intercluster links and the makeup of the clusters
to which it was linked. Once the properties of a cluster
were identified, the properties of those clusters closely
linked to the known cluster could be inferred. For example,
clusters 1 and 12 fall at one end of the distribution;
cluster 1 with one link (with 12) and cluster 12 with two
links (with 1 and 7). Referring to figure 6-3, the graphs
of clusters 1 and 12 are very similar in shape and range.
It was concluded then, that clusters 1 and 12 describe ground
cover conditions that are spectrally similar. Such insight
gained from comparison of the intercluster relationships
evident from both the graphs of the cluster means and the
mean distance tables was used to separate the clusters into
groups with similar spectral characteristics. The cluster
grouping in table VI-5 also indicates those cluster/categories
that are likely to be confused because of spectral
similarities.

TABLE VI-5. — MSS CLUSTER GROUPS
(AUGUST 29, 1972, DATA)

Cluster No.	Nearest Clusters (Within Mean Distance of 4.0)
1	12
12	7, 1
7	3, 4, 12, 6
4	6, 7, 3, 8
8	2, 6, 3, 9, 4
2	8, 6, 9, 3
6	2, 8, 3, 4, 7, 9
3	6, 2, 9, 7, 10, 8, 4
9	2, 3, 10, 8, 6, 11
10	11, 9, 3
11	10, 13, 9
13	11, 5
5	13

The next step in the cluster analysis was to relate the 13 clusters generated by ISOCLS to the land-use scheme described in section 4.2. A gray map produced by the ISOCLS run was carefully scrutinized and hand colored. The patterns that emerged from the colored gray map were then visually correlated with the 1970 HATS land-use map and the manual classification maps produced in the investigation. These maps appear in section 7.

Once a few key features were identified by cluster number, clusters that had similar mean radiance characteristics were identified by examining the cluster curves, the cluster means, and mean distance tables. For example, cluster 7 was seen

to fringe large lakes and its radiance curve (fig. 6-3) was
similar to that of water (clusters 1 and 12) although in bands
6 and 7, cluster 7 appeared to more closely match the vegeta-
tion curves. It became obvious then that cluster 7 was likely
a combination of water and vegetation. The pixels composing
cluster 7 were imagi..g both water (in the lakes) and vegeta-
tion (on the shoreline). The cluster was finally designated
wetland, being a combination of water and vegetation. Large
groups of cluster 7 occurred to the east of Lake Houston.
These were determined to be flooded ricefields after checking
high-altitude aerial photographs of the area. At certain
stages in its vegetative cycle, irrigated rice could be
expected to have a spectral response similar to some swamps
and marshes.

After studying the cluster statistics, gray-map print-
outs, aerial photography, and existing maps of the study
area, each of the 13 clusters was assigned to one of the 5
general land-use categories. From the number of the sample
pixels comprising each cluster, the percentage of occurrence
of clusters in each of the five general land-use categories
was calculated (table VI-6). Two color-coded cluster maps
wer'; produced on the JSC DAS computer by assigning colors
to the clusters. One map (fig. 6-4A) was generated by
aggregating the clusters to represent the two basic families
of cluster curves (vegetated and nonvegetated surfaces) with
water being shown as a separate subgroup. Figure 6-4B was
generated by assigning shades of basic colors to represent
the following subgroups of cluster curves:

1. Water/nonforested wetland — blues.

2. Agriculture/rangeland — yellows.

3. Forest — brown.

4. Urban/built-up — reds.

TABLE VI-6. — ISOCLS CLUSTER ASSIGNMENT AND PERCENTAGE
OF OCCURRENCE

Land-Use Category	Clusters	Occurrence, Percent
Water	1, 12	3.8
Nonforested Wetland	7	0.8
Forest Land	4	42.7
Agriculture/ Rangeland	2, 6, 8	37.8
Urban/Built-up	3, 5, 9, 10, 11, 13	14.9

6.2.2 Supervised Classification

Once each of the 13 ISOCLS clusters was assigned to a
land-use category, the stage was set to perform a classifi-
cation of pixels in the study area using a supervised pattern
recognition algorithm. The maximum likelihood classification
algorithm, LARSYS, which performs a supervised data grouping
was selected for this effort. In a standard approach, the
pixels that comprise areas representative of the land-use
classes to be identified and delineated are input into LARSYS
as training fields and subjected to a statistical analysis.
The statistical parameters of each training field are computed,
with the training field statistics aggregated to yield class
statistics. The statistics represent the average relative
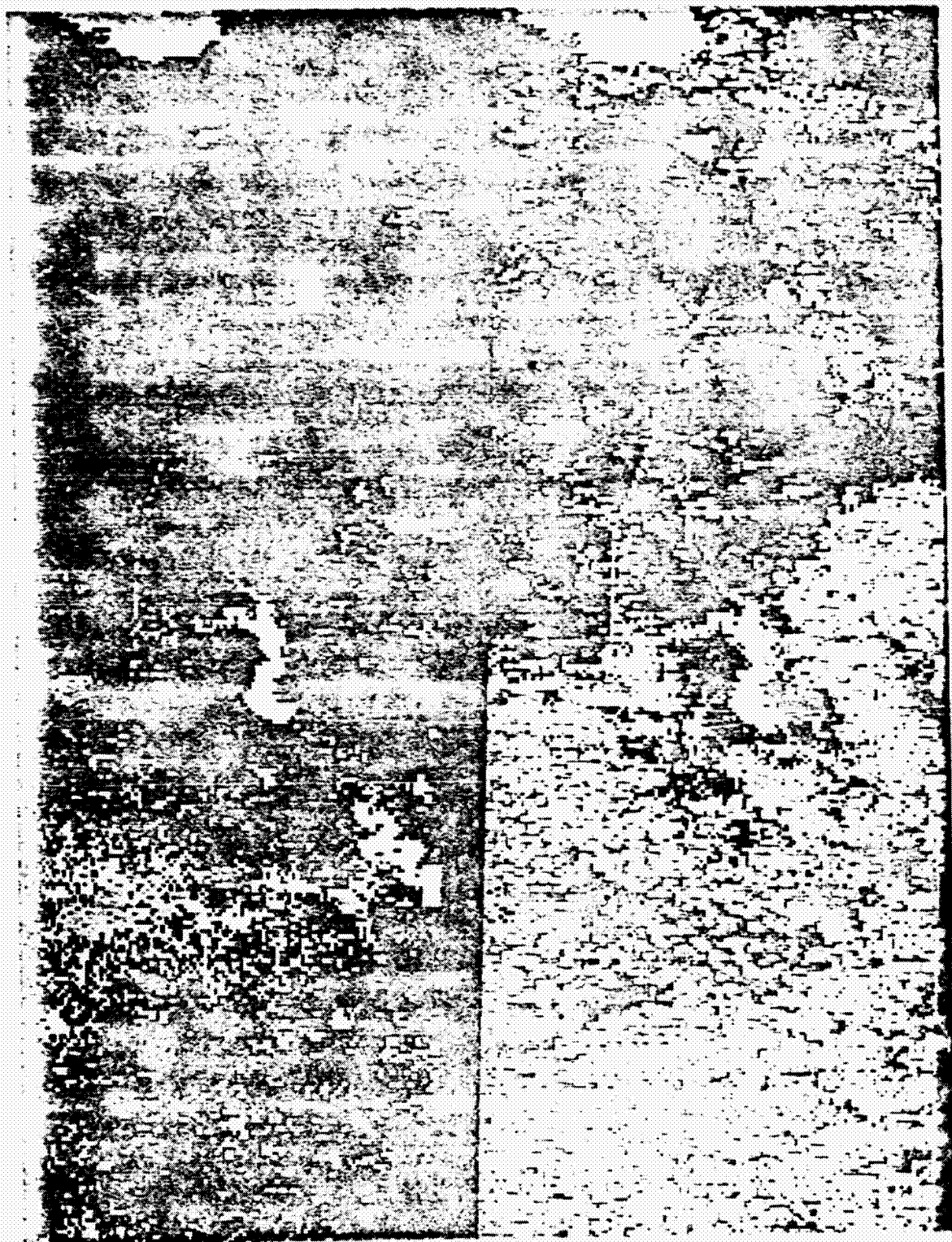
Figure 6-4.— Study area cluster map.

response (or radiance) in each band. With the training
fields representative of each class selected and their
statistics computed, the classification decision is made
at each pixel of the area to be classified.

The nonsupervised approach, ISOCLS, was employed to
produce training clusters. These training clusters
consisted of a variable number of pixels with similar spec-
tral characteristics. The statistics (means and covariance
matrices of the respective clusters generated by ISOCLS)
were substituted for training fields statistics in the
LARSYS-II supervised classification approach. Thus, the
statistics on which the classification of LARSYS-II was
based were those describing spectrally similar pixels
throughout the study area. This approach has the advantage
of eliminating the need of expending effort in selecting
the training fields, which are ordinarily used to generate
the statistics describing a land-use category to be
classified.

Success in substituting training clusters for training
fields encouraged extending the supervised computer
classification procedures to classify every pixel within
the study area rather than classifying just a sample of
pixels. To keep within the computer storage capacity, it
was necessary to divide the study area into seven, equal-
width, north-south strips and process each strip as a
separate computer run. The film output of these runs was
then mosaicked together to form a classification map of the
entire study area.

A Level-I classification map was produced in which each pixel was classified by a color code into one of the following five land-use categories:

1. Urban and built-up.

2. Agriculture/rangeland.

3. Forest land.

4. Water.

5. Nonforested wetland.

The same classification procedures were used to produce another classification map in which the Level-I urban and built-up category was divided into the following Level-II categories:

1. Residential.

2. Commercial/industrial/transportation.

3. Open and other.

Because the spectral responses of the residential and open and other categories were related to the proportion of vegetation in the scene, it was necessary to reassign several of the original agriculture/rangeland clusters (2, 6, and 8) to these Level-II categories. Consequently, to determine whether these clusters were representing Level-I agriculture or Level-II urban features, it was necessary to know the geographic boundaries of the urban area. Clusters 2, 6, and 8 which fell within the known perimeter of the urban areas could then be considered Level-II categories. When they fell outside of the known urban fringe, they could be classified as Level-I categories.

The results of using supervised classification procedures in the study area are reported in section 7.2.

6.3 CLASSIFICATION ASSESSMENT

An assessment of the different approach was performed by measuring the agreement between the classified data and base reference data. Five test sites, ranging in size from 21 km^2 (8 mi^2) to 104 km^2 (40 mi^2), were established in the study area. Base reference data were established by visually classifying land use in each site from high-altitude, infrared Ektachrome photography acquired on April 22, 1972. Each site was divided into 2.6-km^2 (1-mi^2) quadrats, and the percent occurrence of each class in each quadrat was measured using a dot sampling technique. The same procedure was performed on each class in each classification product, except for the computer classification maps, where pixels in each class were counted and converted to percent occurrence. Percent agreement of each classification product with the base reference data on a class-by-class basis was then calculated. The formula for calculating classification assessments is given in appendix A.

A regression analysis was performed to determine how well each classification product served as an estimator of each of the Level-I land-use classes. A linear regression was fit for each product/class versus the base reference data, with the correlation coefficient and the standard error of the estimate providing the indicator of performance relative to the reference base. Analysis of covariance was then performed to determine if there was any significant difference between products as class estimators and to determine which product was the best class estimator.

Details of the statistical procedures used in making
the classification assessments are reported in appendix A.
The actual agreements that had been achieved by the various
data analysis approaches are reported in sections 7.1 and
7.2. The regression analysis using the same data as the
original analysis of the land-use data is found in appendix B.
Appendix C contains the same statistical analysis utilized
in appendix B but presents an alternative scheme for
sampling the original data base.

7.0 <u>RESULTS</u>

7.1 CONVENTIONAL IMAGE INTERPRETATION

Figure 7-1 shows the results of manually grouping the original HATS land-use categories into classes which would be most compatible with the Level-I land-use categories in Circular 671. Delineations of Level-I land-use categories obtained from the October 4, 1972, ERTS-1 imagery over the study area are shown in figure 7-2. A cursory examination of these two figures reveals impressive similarities in category delineations. As might be expected the delineations from the higher resolution aerial photographs were more detailed, but the general patterns and geographic distributions of both delineations were quite similar.

Difficulties were encountered in differentiating agriculture and rangeland categories primarily because of the following reasons:

1. Spatial resolution of the MSS was not sufficient to resolve the small, regular rectilinear field patterns normally associated with the type of agricultural cropping practices found in the study area.

2. Much rangeland in the study area was comprised of grasslands with spectral responses similar to improved pastures and some croplands of the agriculture category.

3. A considerable amount of grazing in the study area is done on brushlands which merge into forest lands; on coastal grasslands which are interspersed with nonforested wetlands; and on natural lands where extractive industries (oil, gas, and sulfur wells) are the dominant economic

N

LEVEL I LAND USE IN HATS

U/B URBAN BUILT UP

WATER

F FOREST

R RANGE

Ag AGRICULTURE

WI NON FORESTED WETLANDS

SOURCE

1 250,000 HATS LAND USE INVENTORY
PREPARED AT NASA MSC

1973
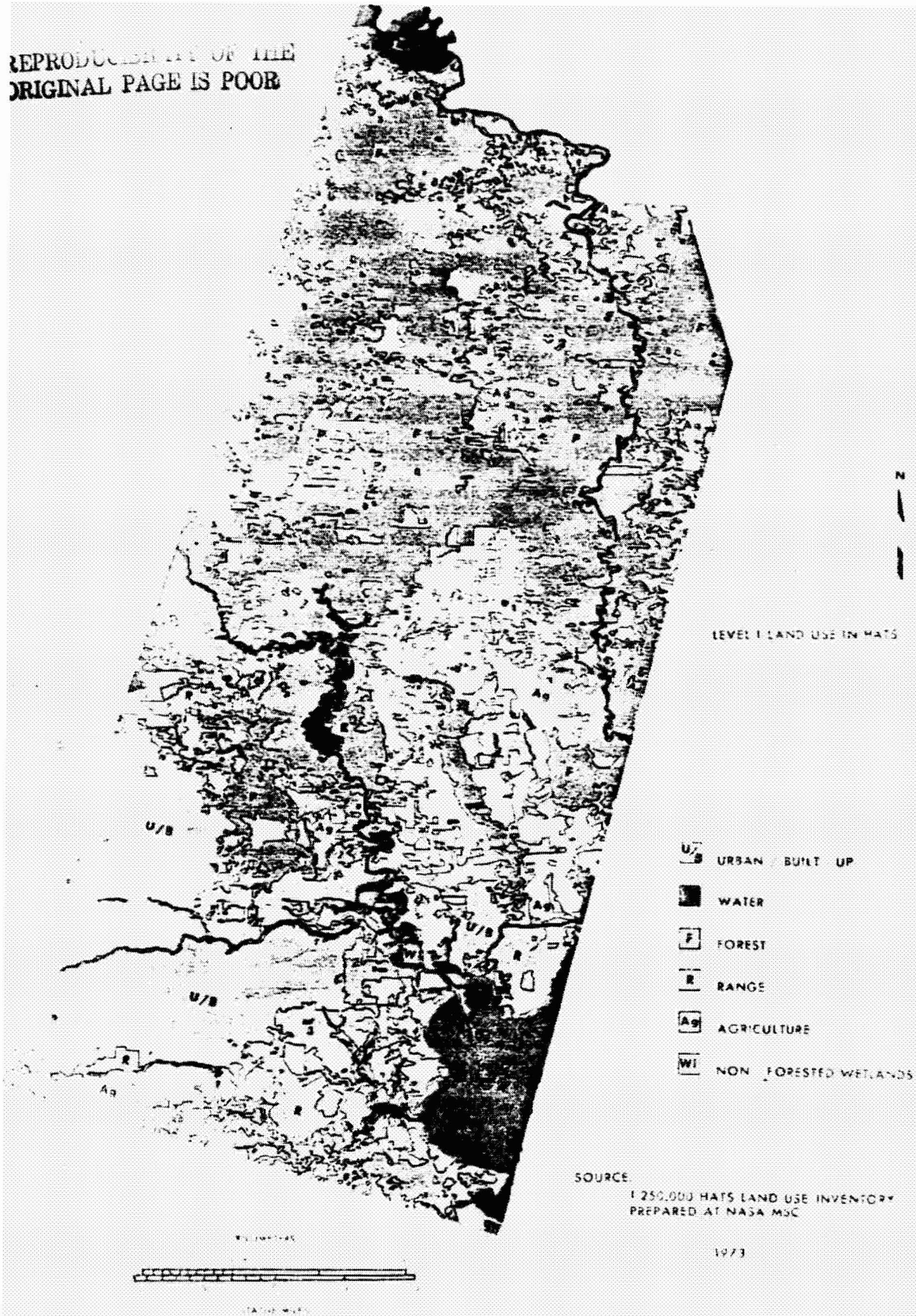
Figure 7-1. - Level - I Land use from aerial photography.

LEVEL 1 LAND USE
INTERPRETATION OF
STUDY AREA IN HATS

Ag

Ag

Cl

A

R

Ag

Wl

Cl

Cl

▓ URBAN/ BUILT-UP

▓ FOREST

Ag AGRICULTURE

▓ WATER

▓ RANGE LAND

Wl NON-FORESTED
WETLANDS

Cl CLOUDS

▓ HIGHLY REFLECTIVE
COMMERCIAL INDUSTRIAL AREAS

SOURCE: 13.5X ENLARGEMENTS OF
70MM ERTS - 1 BULK IMAGERY;
BANDS 5 & 7
ACQUIRED ON 4 OCTOBER 1972

1973

Figure 7-2.- Level-I land use from ERTS-1 imagery.

activities, but where grass or brush vegetation remains the predominant surface cover.

4. Much of the unoccupied land within and along the urban fringe is devoted to grazing. The spectral response of some of these grasslands is indistinguishable from some vegetative surfaces in the urban and built-up lands.

Upon completion of the conventional image interpretation phase of the ERTS-1 imagery, it was deemed advisable because of the above reasons to combine the agriculture and rangeland Level-I categories into one category for all subsequent analyses.

By using the sample test sites and classification assessment procedures described in appendix A, a measure of agreement was determined for the interpretation of both black-and-white ERTS-1 imagery and JSC color composite imagery. A tabulation of the agreements achieved in interpreting these two types of imagery is shown in table VII-1. It can be noted that neither imagery seems to have a distinct advantage over the other imagery, although a slight advantage is evident when interpreting forests with black-and-white imagery or using the color composite imagery for interpreting the agricultural/rangeland category.

For both the forest land and the agriculture/rangeland categories, agreement varied from site to site; and this variation appears to be related to the variations in the percent occurrence of the class. This same relationship was noted in the analysis of the reference base data.

TABLE VII-1. — LEVEL-I AND -II LAND-USE PRODUCT AGREEMENT WITH BASE DATA

(based on percent occurrence)

Land-Use Class	Sample Test Site	Number (100-Point Counts Each)	Base Quadrats		Conventional Image Interpretation				Computer Classification	
					Black and White Imagery		Color Composite Imagery			
			Class Occurrence		Class Occurrence		Class Occurrence		Class Occurrence	
			Count	Percent	Count	Percent	Count	Percent	Count	Percent
	1	33	2427	73.5	2394	72.5	2800	84.8	2383	72.2
Forest	2	38	3319	87.3	3473	91.4	3455	90.9	3340	87.9
	3	15	367	24.5	293	19.5	265	17.7	240	16.0
	4	4	209	52.3	267	66.8	286	71.5	192	48.0
Cumulative Total		90	6322	70.2	5427	71.4	6806	75.6	6155	68.4
	1	17	736	43.3	552	32.5	270	15.9	719	42.3
Agriculture/ Rangeland	2	7	299	42.7	100	14.3	219	31.3	214	30.5
	3	35	3009	86.0	2981	85.2	3245	92.7	2825	80.7
Cumulative Total		59	4044	68.5	3633	61.6	3734	63.3	3758	63.7
Water	4	6	325	54.2	259	43.2	247	41.2	296	49.3
Level-I Urban	5	39* _11**_	3724* _1035**_	95.5* _1035**_	1100**	100.0**	3855	98.8	1642	42.1
Level-II Urban		19	1293	68.1	Residential		1750	92.1	1514	79.7
		8	137	17.1	Commercial/ Industrial/ Transportation		60	7.5	82	10.3
		12	200	16.7	Open and Other***		***	***	124	10.3

*August 29, 1972, ERTS-1 data used for all analyses except**.

**October 4, 1972, ERTS-1 data base covered only part of test site in black-and-white imagery.

***Category not delineated by conventional image interpretation.

Because no attempt was made to interpret Level-II urban
categories on the black-and-white imagery, no direct
comparison could be made with the Level-II categories inter-
preted from the color composite imagery. It is believed
that the lower level of agreement obtained in interpreting
Level-I urban categories on the black-and-white imagery may
be due primarily to the smaller sample that was available
with the October 4 imagery in which an orbit shift resulted
in only 11 quadrats being covered in the urban test site.
The fact that the base data (aerial photographs) were not
acquired concurrently with the ERTS-1 data (April 22 versus
August 29) may have contributed to a level of agreement
below those expected in the agriculture/rangeland category
due to seasonal changes in vegetative cover. By comparing
the actual class counts of the categories in each test site,
it will be noted that counts in the forest category are
likely to be overestimated, whereas those in the agriculture/
rangeland category are likely to be underestimated. This
appeared to be more often the case with the color composite
imagery than with the black-and-white imagery. The Level-I
urban category was overestimated by a relatively small amount
when interpreting both types of imagery. A more noticeable
error in underestimating occurred in interpreting the water
category on both types of imagery.

It is believed that the low accuracy in interpreting water
can be attributed to one or more of the following reasons:

1. Water had a relatively low percentage of occurrence
in the study area.

2. Spectral response of water surfaces could vary
greatly because of variations in Sun angle or in levels of
turbidity.

3. Flooded ricefields which should be in the agri-
culture/rangeland category were often confused spectrally
with ponds and small lakes.

It is believed that the percent agreement (43.9 percent)
of the commercial/industrial/transportation category is not
representative of the actual accuracy with which this
category can be mapped from the ERTS imagery by conventional
image interpretation. It is evident from figure 6-1 or 6-2
that many transportation lines (highways, utility lines, etc.)
that are well below the resolution limit of the scanner still
appear as definite linear features on the imagery. These
linear features may not be shown on the imagery as continuous,
solid lines because of the scanning characteristics of the
sensor and the particular orientation of the feature in
relation to the satellite orbit. Thus, it is suspected
that the point-grid method of sampling the imagery may not
always represent the true count of pixels coincident with
the linear features.

7.2 COMPUTER CLASSIFICATION

Film output maps from the supervised computer classifi-
cations of land use in the study area are shown in figures 7-3
and 7-4. The basic difference between these two maps
was the manner in which the Level-I urban category in
figure 7-3 was divided into Level-II urban categories
(residential, commercial/industrial/transportation, open,
and other). The land-use patterns outside of the urban
areas remain virtually the same on both maps. A comparison
of these two maps with the land-use maps obtained from the
interpretation of aerial photography (fig. 7-1) and ERTS-1

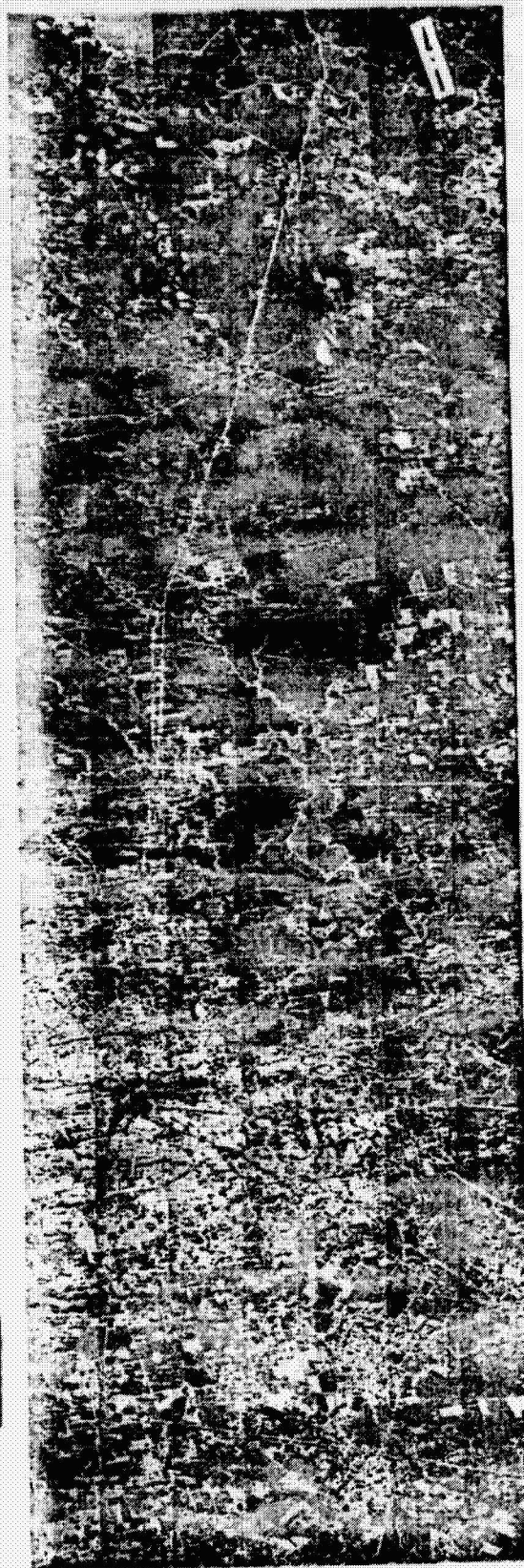Figure 7-3.— Level-I computer classification of study area.

Figure 7-4.— Level-II computer classificaton of study area.

imagery (fig. 7-2) reveals impressive similarities of land-use delineations in the rural areas. The delineations on the manually interpreted maps appear somewhat more generalized than those on the computer land-use maps. This was not unexpected because of the relatively small scales (1:250,000 and 1:120,000) used in compiling the manually interpreted maps. Details smaller than those actually drawn on these maps were discernible on the original imagery, but there were physical limitations to the size of the delineations which could be drawn by hand on the small-scale overlays. In contrast, it was possible to depict details on the computer maps which had dimensions of only one pixel (approximately 1.1 acres or 0.45 ha).

Major differences between the computer maps and the manually interpreted maps are most noticeable in the delineations of the urban areas. It is evident that the areal extent of metropolitan Houston is much greater on figures 7-1 and 7-2 than on figures 7-3 and 7-4. This would indicate that the urban fringe, where a transition from predominantly vegetative surfaces to predominantly paved or bare surfaces occurs, is an area without distinct or unique spectral characteristics. Consequently, the computer classifications could not distinguish between trees or grass located in urban areas from trees or grass found in rural areas. On the other hand, the human interpreter could delineate these areas more readily by using spatial characteristics (size, shape, location) with tones or colors and textures as recorded on aerial photographs or ERTS-1 imagery.

As discussed in section 6.2.2, spectrally similar pixels were grouped together into clusters and assigned to either

agriculture/rangeland or Level-II urban categories depending
on whether the entire study area was being classified or
whether only the metropolitan Houston area was being
classified. This made it possible to assign Level-II urban
categories to groups of pixels within the urban complex
which normally would have been classified into the Level-I
agriculture/rangeland cluster. Consequently, greater
classification accuracies could be achieved by selecting a
sample test area that was entirely within the concentrated
urban area rather than one that straddled the urban fringe
where confusion with agriculture/rangeland would occur.

The agreements achieved by computer classification
procedures are shown in table VII-1. It should be noted that
the lowest agreement occurred in classifying the Level-I
urban category. This collaborates the classification
confusion which was apparent in comparing the delineations
of urban and agriculture land use in figures 7-1 and 7-4.
The greatest classification agreements were achieved in the
forest category. This was probably the result of the large
expanses of forest and the relatively homogeneous spectral
response of most of the forest cover.

By comparing the computer classification agreements
with the conventional image interpretation agreement, it
will be noted that only minor differences in agreement
occurred in the forest and agriculture/rangeland categories.
Considerably better agreements were achieved in classifying
water by computer classifications than by interpreting
either black-and-white or color composite imagery. This
appeared somewhat antithetical, because it was expected that
the same reasons cited for the relatively low agreements

achieved in interpreting water from black-and-white or color
composite imagery (section 7.1) would also apply to the
computer classification procedures. The fact that bodies
of water as small as one pixel (1.1 acre or 0.45 ha) in
size could be classified by the computer was probably a major
reason why better agreements were achieved by computer
classification procedures. However, it is believed appro-
priate to consider this finding only tentative, because the
total number of quadrats containing water within the study
area provided a relatively small statistical sample. It
should be noted that in only two instances (forest site
number 2 and residential) did the computer overestimate the
number of class counts. The difference in the forest class
count is considered insignificant, but the difference in
residential class count is considered important because it
emphasizes the complex nature of the residential category
in which vegetation tends to confuse the category with non-
urban categories.

An attempt was made to identify several major classifi-
cation anomalies which resulted from spectral similarity of
different classes. Some areas of high radiance (bare soil,
recently cutover forest, stubble from recently harvested
rice, etc.) were misclassified as urban/built-up category.
Irrigated ricefields were sometimes classified as non-
forested wetlands. Vegetated sections of metropolitan
Houston frequently were classified as agriculture/rangeland.

8.0 CONCLUSIONS

This investigation demonstrated that it was feasible to
use computer classification techniques to classify Level-I
land use over a relatively large area from ERTS-1 MSS data.
The key to this success was the use of a data sampling
technique in conjunction with the nonsupervised clustering
algorithm ISOCLS in stage I of the two-stage computer
classification approach adopted for this investigation. A
small (3 percent) sample of the available digital data was
sufficient to identify the basic spectral variations asso-
ciated with Level-I land-use classes throughout the study
area. The first stage of this classification procedure
required less than 15 min of computer processing time.

In the second stage, class statistics generated in
stage I were utilized as input to the maximum likelihood
classification algorithm LARSYS to classify all data points
in the ground scene.

The approach used in this study differed significantly
from standard pattern recognition procedures, which require
establishing ground truth training fields for each class in
the scene, developing class signatures, and extending these
signatures to classify the total area under consideration.
In the standard approach, the larger an area to be classi-
fied, the larger the number of training fields that would
be required.

Although the computer classification approach offered
good potential for classifying Level-I land use over large
areas, there also appeared to be some conditions under which

conventional image interpretation procedures could be used
to advantage. Where spectrally homogeneous features pre-
dominated (e.g., forest land) computer classifications
achieved high levels of agreement. However, computer
classification agreement decreased where features were
spectrally heterogeneous and spatially complex (e.g., urban
areas). Under these conditions it was advantageous to use
conventional image interpretation procedures to either
aggregate some computer classifications into desired spatial
patterns, or to preclassify specific areas with similar
spatial characteristics so that separate computer classi-
fications could be made for each specific area.

Finer details were displayed on the computer classi-
fication products than on the products obtained by conventional
interpretation of ERTS imagery. This was because the computer
classified each individual pixel and the output display was,
therefore, not affected by the scale of the original data,
as was the case where delineations were made manually on
the ERTS imagery. Despite the scale limitations of the
ERTS imagery, conventional image interpretation techniques
offer a valid and economical method of classifying large
areas into Level-I and some Level-II land-use categories,
particularly in those instances where sophisticated computer
processing facilities are not available. One distinct
advantage of this method is that the interpreter can utilize
spatial pattern recognition as well as a nominal amount of
spectral discrimination in interpreting the ERTS imagery.

9.0 RECOMMENDATIONS

9.1 SUGGESTIONS FOR IMPROVING CLASSIFICATION METHODOLOGY

On any future manual classification of ERTS MSS imagery,
consideration should be given to determining the feasibility
of using the 9- by 9-inch color composite imagery. Enlarging
and rectifying this imagery to a scale of 1:250,000 is
feasible, and it may prove to be a reliable data source for
mapping land use at a usable scale.

To improve on the accuracy of signature extension,
ground truth sample sites should be selected from throughout
the area being classified. Ground truth surveys and aircraft
underflights in agriculture areas should be as close to
synchronous with the satellite overpass as possible.

The process of producing the supervised land-use
classification maps was operationally cumbersome and in many
instances inefficient. Ideally one would like to have a
single film transparency produced from each classified data
tape, which would image the 96- by 25-mi ground swath
on a single 9-inch wide transparency which would be distor-
tion free. This film transparency would then be rectified
and reduced to a negative film clip. Prints for mosaicking
(at any scale), viewgraphs, etc., could then be made when
desired.

To achieve film outputs to these specifications requires
changes and improvements in the LARSYS pattern-recognition
algorithms and improvements in the DAS film recorder. The

LARSYS-II processing must be speeded up and a method
developed for rapidly classifying all pixels and all scan
lines on the ERTS-digital, scene-corrected magnetic tape.
Output should be on a single digital tape compatible with
the DAS, so that a continuous film strip of the entire
scene could be produced. A higher resolution film recorder
may be in order to handle these classification tapes. The
most critical need in the film recorder is the elimination
of distortions in image size that consistently occur between
film recording runs.

In order to improve the accuracy of land-use deter-
minations via computerized classification, it is recommended
that models be developed which would incorporate spatial
relationships into a classification algorithm. For example,
in the Level-I land-use analysis performed in this investiga-
tion, clusters 2, 6, and 8, as generated by ISOCLS, described
distributions of spectral signatures that represented
agriculture/range as well as vegetation within urban areas.
Assigning a pixel to the agriculture/range class or urban
class is necessarily a matter of spatial interpretation.
If the pixel in question is extensively surrounded by
clusters describing vegetation, it would seem to be in an
agriculture/range or forested area. However, if the pixel
is in the midst of a heterogeneous group of pixels, including
pixels of high reflectance as vegetative clusters, it
might suggest an urban/built-up area. The greater the
spectral heterogeneity, the more likely the area is urban/
built-up.

A spatial dimension could be employed if a grid describing quadrats was superimposed over an area. The occurrence of the various clusters within each quadrat could thus be counted, as an option within the classify module of LARSYS. The classification of pixels assigned to questionable clusters would then be dependent on the distribution of cluster counts within that quadrat.

In order to improve the efficiency of the supervised classification algorithm used in this study and to thereby facilitate the classification of much larger areas, a number of data-sampling procedures should be investigated. First the feasibility of using less than all four bands of data should be evaluated. All two- and three-band combinations should be tested over controlled test sites. An accuracy analysis of the results should then be conducted to objectively assess the relative accuracy of each of the combinations tested. This investigation should also entail sampling different combinations of lines and pixels to determine the least number of data points required to accurately classify a given scene. The goal should be to classify very large areas with a minimum of data points required for inputs.

9.2 MODIFICATION OF THE USGS LAND-USE SCHEME FOR UTILIZATION OF ERTS DATA

The results of this investigation suggest that the USGS hierarchy may need to be modified on a spectral basis to render it more useful in automatically classifying land use from ERTS-type data. In particular, the urban and built-up Level-II categories should be consolidated into spectrally similar groups.

It may be necessary to consolidate certain Level-I
categories based on local conditions. In southeast Texas,
the site of this investigation, rangeland could not be
spectrally differentiated from cropland and pasture. As
a result it was necessary to combine rangeland with agricul-
tu...g. Rangeland in west Texas and other semiarid areas
may be more readily delineated from agricultural land.

Further investigation is needed to resolve the diffi-
culties encountered in differentiating Level-II categories
of forest, water, wetland, and agriculture classes.

10.0 BIBLIOGRAPHY

Aldrich, R. C.: Classifying Forest and Nonforest Land on
Space Photography. Pacific Southwest Forest and Range
Experiment Station, Forest Service, USDA (Berkeley,
California), Mar. 1969.

Aldrich, R. C.: Space Photos for Land Use and Forestry.
Photogrammetric Engineering, vol. XXXVIII, no. 4,
Apr. 1971, pp. 389-401.

Anderson, J. R.; et al.: A Land Use Classification System
for Use with Remote Sensor Data. Geological Survey
Circular 671, USGS (Washington, D.C.), 1972.

Anderson, J. R.: Land Use Classification Schemes. Photo-
grammetric Engineering, vol. XXXVII, 1971, pp. 379-381.

Laboratory for Agricultural Remote Sensing: Remote Multi-
spectral Sensing in Agriculture. Research Bulletin,
vol. 3, no. 844, Purdue University (Lafayette, Indiana),
Sept. 1968.

MacDonald, R. B.: The Application of Automatic Recognition
Techniques in the Apollo 9 SO65 Experiment. Third Annual
Earth Resources Program Review, NASA JSC (Houston, Texas),
Dec. 1970.

Marble, Duane; and Horton, F.: Remote Sensing and the Study
and Planning of Urban Areas. Earth Resources Applications
Program Status Review; Geology, Geography, and Sensor
Studies, vol. 1, NASA JSC (Houston, Texas), Sept. 1968,
pp. 7-1 — 7-B-2.

Minter, R. T.: ISOCLS, Iterative Self-Organizing Clustering
Program, Program CO94. NASA JSC (Houston, Texas),
Contract NAS 9-12200, Oct. 1972.

Nunnally, N. R.; and Witmer, R. E.: A Land Use Interpreta-
tion Experiment. Commission on Geographic Applications of
Remote Sensors, Association of American Geographers.
Technical Report 69-5, Jan. 1970.

Pettinger, L. R.; et al.: Multiple Resource Inventory on
   Space and High-Altitude Photography. Second Annual Earth
   Resources Aircraft Program Status Review, NASA JSC
   (Houston, Texas), Sept. 1969.

Paulton, Charles E.: Range Resource Inventory from Space
   and Supporting Aircraft Photography. Second Annual Earth
   Resources Aircraft Program Status Review, NASA JSC
   (Houston, Texas), Sept. 1969.

Rudd, R. D.: Macro Land Use Mapping with Simulated Space
   Photos. Photogrammetric Engineering, vol. XXXVII, 1971,
   pp. 365-372.

Steiner, Dieter: Use of Air Photos for Interpretating
   and Mapping Rural Land Use in the United States. Photo-
   grammetria, vol. XX, 1965, pp. 65-80.

Thrower, N. J. W.: Land Use in the Southwestern United States --
   From Gemini and Apollo Imagery. Annals Map Supplement
   Number Twelve, Annals Association of American Geographers,
   vol. 60, nrl, Mar. 1970, pp. 208-209.

Ulliman, J. J.: The Use and Cost of Air Photos in Land Use
   Planning and Classification. Minnesota Report 98, Forestry
   Series no. 7, University of Minnesota Agriculture Experi-
   ment Station.

_____: Thematic Land Use Mapping with Spacecraft
   Photography and Radar. Earth Resources Applications Program
   Status Review; vol. 1, Geography, Geology, and Sensor
   Studies, NASA (Houston, Texas), 1968, pp. 8-1 — 8A-4.

APPENDIX A

ACCURACY ANALYSIS STATISTICAL PROCEDURES

ACCURACY ANALYSIS STATISTICAL PROCEDURES

An objective statistical method was developed in the course of this investigation to measure and assess the relative merits of each classification product obtained in the analysis approaches. A ground truth reference base was developed from small-scale (1:120,000) color infrared aerial photography obtained over the study area in April 1972.

Because it was impractical to analyze every pixel of data in the entire study area, a sampling procedure had to be considered. Initially, a random sampling procedure was considered; but because of the difficulty in locating data points on the classification products and the reference photography, it was deemed advisable to select five representati·e base reference sites in lieu of a random sampling procedure. These sample study sites were representative in that each contained a preponderance of one or two Level-I land-use categories which were not randomly distributed throughout the entire study area.

The base reference sites are:

Number 1 - a combination forest, agriculture/range, and urban area near Cleveland in the northern part of the study area (fig. A-1).

Number 2 - a predominantly forested area located north and east of Lake Houston (fig. A-2).

Figure A-1.— Cleveland, forest — agriculture/range —
urban base reference site no. 1.

Figure A-2.— Forest base reference site no. 2.

Number 3 - an agricultural area east of Lake Houston
        (fig. A-3).

Number 4 - a forest and water area at the northern end of
        Lake Houston (fig. A-4).

Number 5 - an urban area located in the north central part
        of metropolitan Houston (fig. A-5).

Because of the difficulty in locating the same data point
on all three products with any degree of precision, the
selection of a sample population was restricted to that of
utilizing the five sample sites; thus, 20 data points had
to be located (the four corners of the five test sites)
instead of a multitude of sample points.

In order to insure that the reference data possessed
a sufficient level of accuracy and to reduce possible
interpreter bias, two independent interpretations were made.
The results of the two interpretations were then compared
for each class within each quadrat. Discrepancies between
the interpretations were measured using the following
equation:

$$D_{ab} = \left[1 - \left(\frac{|A_{ab} - B_{ab}|}{B_{ab}}\right)\right]100$$

where

D_{ab} = the measure of the agreement between the two
    interpretations for class a in quadrat b expressed
    as a percentage.

Figure A-3.— Lake Houston east, agricultural base
reference site no. 3.
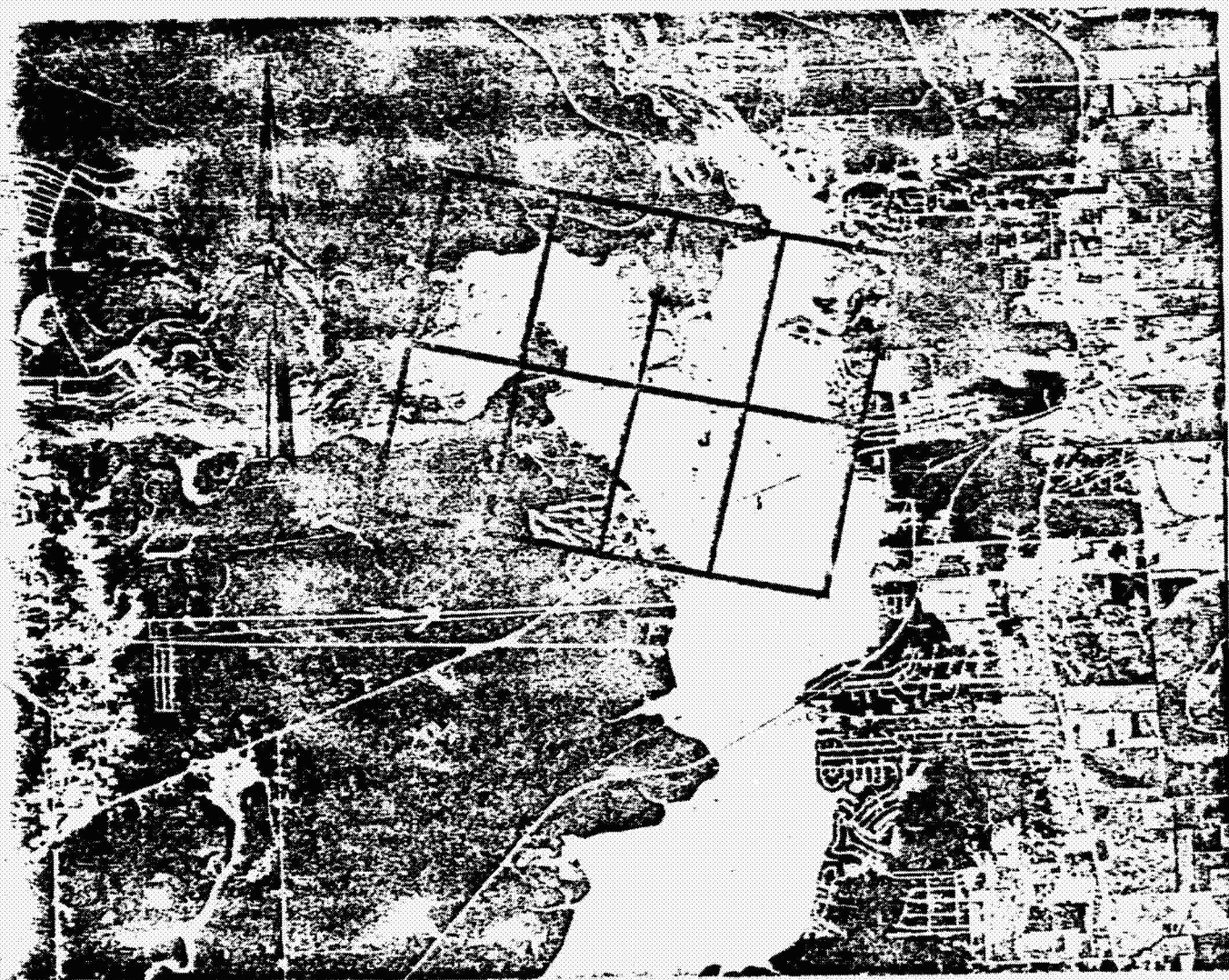
Figure A-4.- Lake Houston north, water base reference site no. 4.

Figure A-5.— Houston, urban base reference site no. 5.

B_{ab} = percent occurrence of class a in quadrat b
as determined by the first interpretation.

A_{ab} = percent occurrence of class a in quadrat b
as determined by the second interpretation.

To strengthen the reliability of the base reference
data, only those quadrats in which a high degree of agree-
ment for a given class was achieved were selected for use
in the accuracy analysis. The threshold for deleting data
was set at 85-percent agreement. Therefore for each class,
only those base reference quadrats with agreements of
85 percent or higher were utilized in the subsequent analysis.
As a result, the wetland class, which is practically non-
existent in the sample sites as well as in the study area,
was eliminated from the accuracy analysis. Tables A-1 and
A-2 show the final baseline data selected for the accuracy
analysis.

In order to acquire an adequate number of data points,
the sample sites had to be relatively large. All but the
Lake Houston north site are approximately 8 mi (13 km)
by 5 mi (8 km) in dimension. The boundaries are rectangular
and coincide with the scan lines and pixel lines in the ERTS-1
data. Each site is 200 pixels wide by 125 scan lines on a
side. Lake Houston north is 100 pixels by 50 scan lines in
dimensions.

The sites were first delineated on a JSC color composite
of the study area which had a grid showing every 50th scan
line and every 50th pixel of the ERTS data. These boundaries

TABLE A-1.— BASELINE DATA SELECTED FOR ACCURACY ANALYSIS

Sites: Class	Cleveland A^1	Cleveland B^2	Forest A	Forest B	L. Houston East A	L. Houston East B	L. Houston North A	L. Houston North B	Urban A	Urban B
Forest	33	73.5	38	87.3	15	24.4	4	52.2	—	—
Agriculture/ Range	17	43.3	7	42.7	35	86.0	—	—	—	—
Urban	—	—	—	—	—	—	—	—	39	95.5
Water	—	—	—	—	—	—	6	54.2	—	—

^{1}A = Quadrat count.

^{2}B = Percent occurrence.

TABLE A-2.— SUMMARY OF BASELINE DATA SELECTED FOR ACCURACY ANALYSIS

Level I Class	No. of Quadrats	Point Count All Quadrats	Total Class Points	Occurrence, Percent
Forest	90	9000	6322	70.2
Agriculture/ Range	59	5900	4044	68.5
Urban	39	3900	3724	95.5
Water	6	600	325	54.2
Level II				
Residential	19	1900	1293	68.1
Commercial/ Industrial/ Transportation	8	800	137	17.1
Open and Other	12	1200	200	16.7

were then transferred to the aerial photography selected as
the base reference (color infrared transparencies).

An 8×5 grid dividing the sites into 40 quadrats was
described on the film. Each quadrat was an area 25 pixels
by 25 scan lines in the ERTS data or 625 pixels in each
quadrat. The water site (Lake Houston north), being smaller,
had an 4×2 grid or eight quadrats.

A 100-point grid (10×10) was constructed to overlay each
quadrat so a 100-point sample could be taken from each
quadrat. A dot-grid of 625 points would have been more
desirable; but because of scale limitations in both the aircraft
data and the manually interpreted ERTS data, the 100-point
grid proved to be a practical compromise.

Land use in the sample sites was determined by over-
laying each quadrat with the 100-point grid and interpreting
the Level-I land-use class at each point. Level-II classes
were also interpreted over the urban site. The number of
points in each class was tabulated and converted to a
percentage.

Although excellent agreement of classification occurred
between the two interpreters, sufficient time and personnel
resources were not available to conduct detailed ground truth.
Surveys to determine what percentage of the agreement may
have resulted from actual errors in land-use interpretations
were made by both image interpreters. Considering the skill
of the two interpreters and the relatively few classes of
land use being interpreted, it was believed most unlikely that

both interpreters would commit the same error in interpreting
the same feature. Thus, a high measure of agreement in
interpretation was most likely agreement in correctness of
interpretation rather than in error of interpretation.

The accuracy calculations for classification products
were computed in the same way the accuracy of the base
reference figures were computed. The accuracies for each
classification product were determined using the following
formula:

$$X = \left[1 - \left(\frac{\left| \sum_{i=1}^{n} A_i - \sum_{i=1}^{n} B_i \right|}{\sum_{i=1}^{n} B_i} \right) \right] 100$$

where

X = percent accuracy for each land-use class.

A = class occurrence (percent) as mapped in each
quadrat from ERTS imagery.

B = class occurrence (percent) in base reference
quadrats.

n = number of quadrats.

APPENDIX B

STATISTICAL ANALYSIS OF LAND-USE DATA

STATISTICAL ANALYSIS OF LAND-USE DATA

The statistical analyses presented in this appendix utilized the same reference data quadrats used in the classification assessment discussed in section 6.3 and appendix A. Utilizing the same data, a more rigorous statistical analysis is demonstrated.

1. Evaluation of the Data Base

The initial data base was composed of 168 quadrats, with
the percent occurrence of a feature distributed according
to the following table:

TABLE B-1.— DISTRIBUTION OF CLASS OCCURRENCE FOR
ORIGINAL QUADRATS

Occurrence of a Feature in a Quadrat, Percent	Number of Quadrats Per Class			
	Forest	Agriculture/Range	Urban	Water
80-100	50	32	41	1
60-79	14	8	0	1
40-59	15	13	0	4
20-39	14	13	0	2
1-19	57	41	35	54
0	18	61	92	106
Total	168	168	168	168

These data were obtained by one photointerpreter evalu-
ating 1:120,000 scale color Ektachrome aerial photography of the
test sites. A second photointerpreter evaluated the same
imagery; and based on the following equation, a quadrat was
either retained for comparison or discarded:

$$\left(1 - \frac{|(A - B)|}{B}\right) * 100 \geq 85 - \text{retain}$$
$$< 85 - \text{discard},$$

where A is the percent occurrence according to the f:` t
interpreter, and B is the percent occurrence according to
the second interpreter.

This technique yielded the following table of quadrats
which are based toward the high percent occurrence quadrats.

TABLE B-2.— DISTRIBUTION OF CLASS OCCURRENCE FOR
ACCEPTED QUADRATS

Occurrence of a Feature in a Quadrat, Percent	Number of Quadrats Per Class			
	Forest	Agriculture/Range	Urban	Water
80-100	50	32	39	1
60-79	13	7	0	0
40-59	10	5	0	4
20-39	4	9	0	1
0-19	12	5	0	0
Total	89	58	39	6

2. Land-Use Class/Product Agreement

The class/product agreements shown in table VII-1 were calculated using the following formula:

$$
X = \left[1 - \left(\frac{\left| \sum_{i=1}^{n} A_i - \sum_{i=1}^{n} B_i \right|}{\sum_{i=1}^{n} B_i} \right) \cdot 100 \right]
$$

where,

X = percent agreement for each land-use class.

A = class occurrence (percent) as mapped in each quadrat for ERTS imagery.

B = class occurrence (percent) in base reference quadrats.

n = number of quadrats retained for each land-use class.

3. Regression on Data From Retained Quadrats

An example of the effect of the retained data used in a regression (least-squares analysis) can be seen in figure B-1. The high concentration of points in the 100-percent region results in that portion of the data having a larger influence on the slope and intercept of the regression curve.

The dashed line is the standard error (66-percent confidence interval) of the curve. Although a detailed analysis of this error was not done, it is obvious that the error is inversely related to the estimate and would be larger in the area of small percent occurrences. This is due to the smaller number of samples that occurs in the region of the curve.

A regression analysis was performed for all products, as well as all classes. The results of this analysis are displayed in table B-3. The biasing of the data towards the higher occurrence quadrats affected all the results of the regression analysis to a greater or lesser extent than the forest land/computer product example.
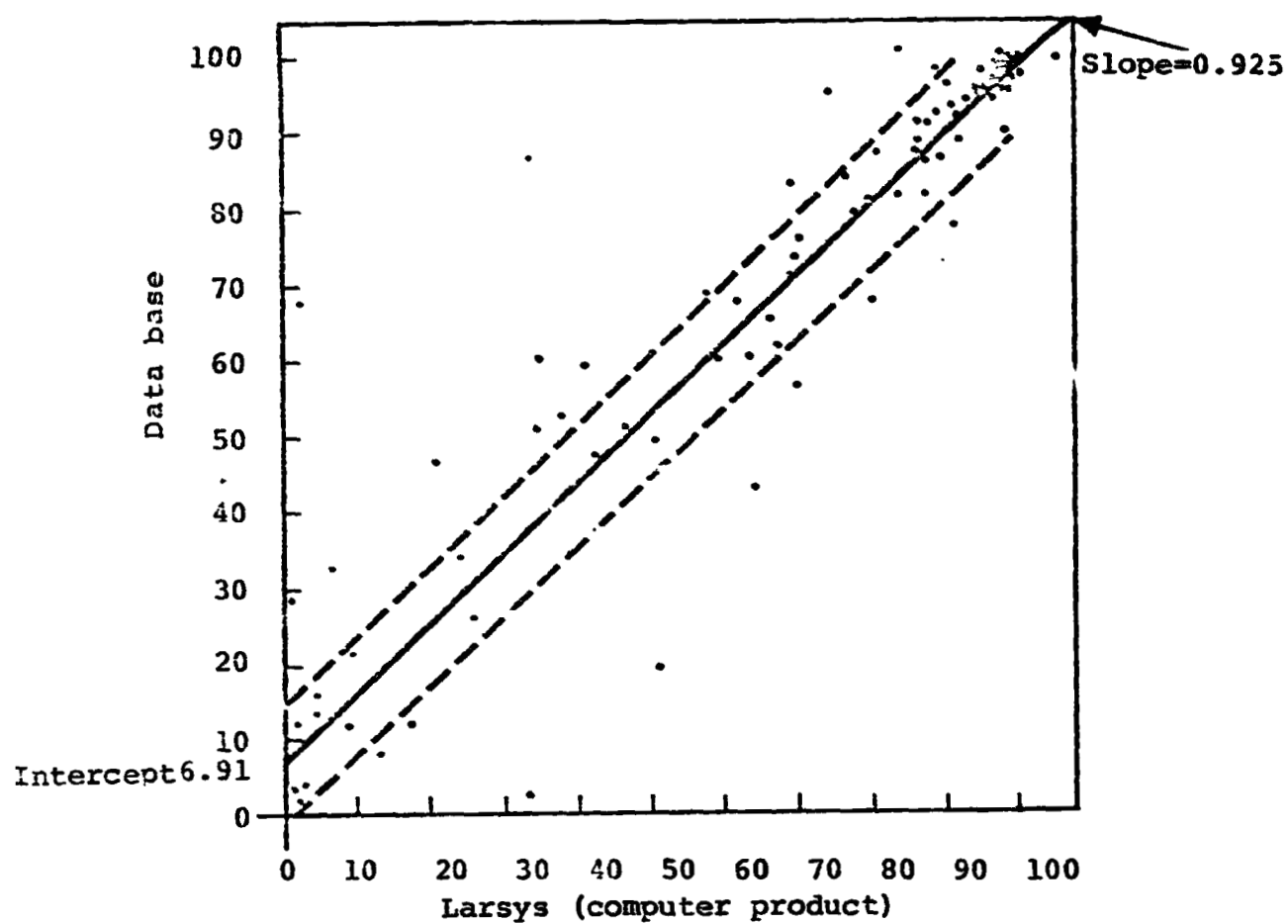
Figure B-1.— Simple regression cf forest land occurrence
from the data base on the forest land occurrence indi-
cated by the computer product for the acceptable forest
land quadrats.

	Black and White Imagery			JSC Color Composite			Computer		
	Equation	Standard Error	Significance	Equation	Standard Error	Significance	Equation	Standard Error	Significance
Forest	$1.8+0.73*E$	19.4	1% $R=0.787$ $R^2=0.619$	$5.8+0.84*E$	14.1	1% $R=0.893$ $R^2=0.797$	$6.9+0.95*E$	9.1	>1% $R=0.953$ $R^2=0.908$
Agriculture/ Range	$9.8+0.79*E$	15.4	1% $R=0.854$ $R^2=0.729$	$24.3+0.67*E$	15.8	1% $R=0.859$ $R^2=0.738$	$6.0+0.98*E$	13.6	<1% $R=0.899$ $R^2=0.808$
Urban	$-9.9*10^{93}+1.9$ $*10^{90}*E$	$9.9*10^{48}$	$R=$	$84.3+0.12*E$	4.1	$R=0.13$ $R^2=0.016_9$	$95.4+0.01*E$	4.1	$R=0.020$ $R^2=00000_6$
Water	$29.3+0.57*E$	15.4	10% $R=0.766$ $R^2=0.587$	$2.7+0.65*E$	19.0	$R=0.$ $R^2=0.368$			>1% $R=0.930$ $R^2=0.805$

NOTE: E = estimate obtained from product; R = correlation coefficient; R^2 = percent reduction error due to regression.

4. Analysis of Covariance

The purpose of the analysis of covariance is to determine if all three products for a class have the same regression equation. Figure B-2 is an example of the three curves for the forest category. The test will determine if the three curves are statistically similar based upon the explained error of each curve.

The process involves determining if all three slopes (one for each curve) are the same. If the slopes are the same, the next step is to determine if the levels of the three curves are the same. If both conditions are met, the three curves are considered to be equal; and one curve is generated for the data base versus all three products. This situation occurred only for the water class. (See table B-4.)

If either of the two tests failed, the two best products were further analyzed by performing a t-test of correlation coefficients; because although there is a difference between the products, none of these has been proved to be superior.
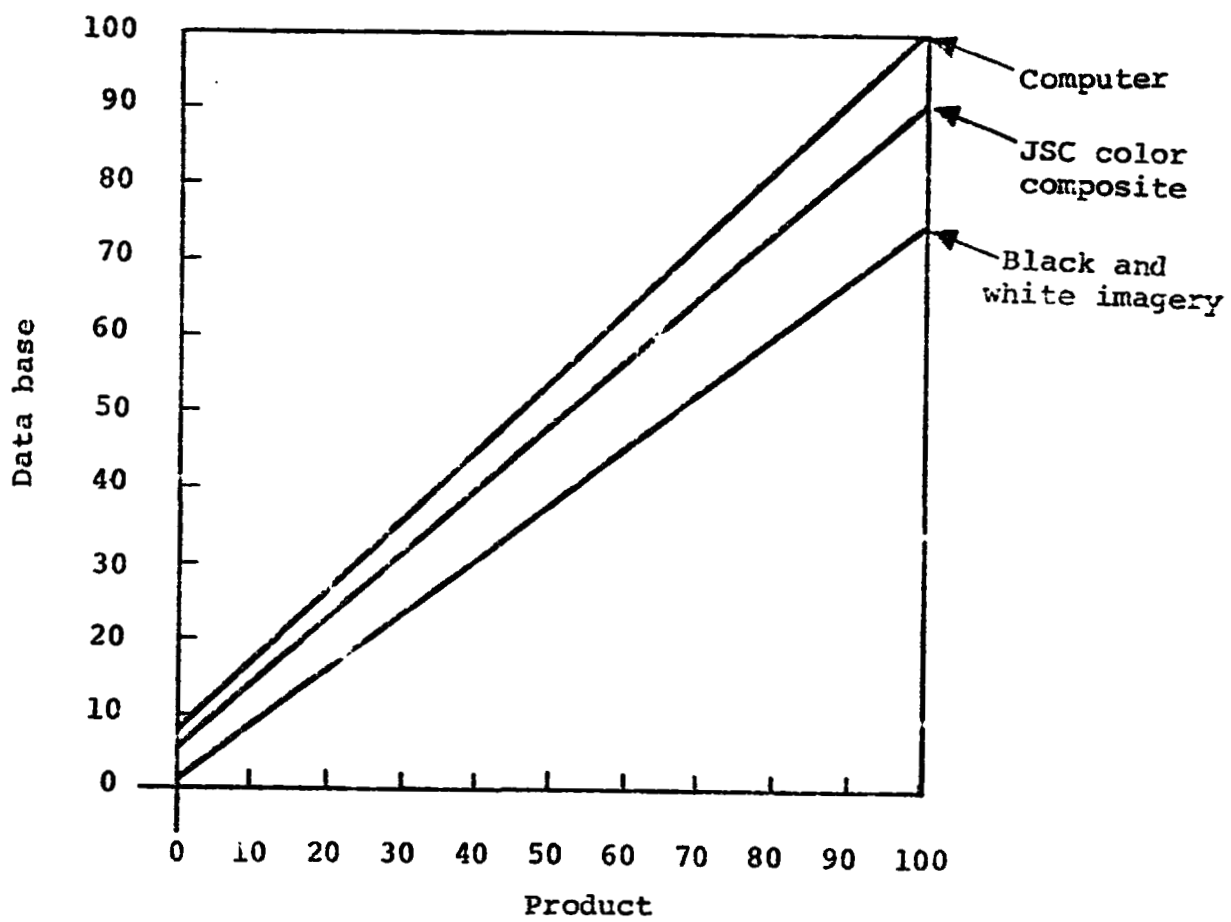
Figure B-2.— The curves resulting from regression of the three forest products on the data base.

TABLE B-4.— RESULTS OF ANALYSIS OF COVARIANCE FOR THE
THREE PRODUCTS FOR EACH CLASS

Class	Level of Significance* Indicated By Test of Slopes, Percent	Conclusion
Forest	>5	Difference exists
Agriculture/ Range	0.1	Difference exists
Water	<10	Insignificant - no difference; test for levels was also <10%; do regression for all products versus the data base
Urban	No tests since no regression curves were significant.	

*Interpretation of "significance" — the evaluation of
the slopes is performed with the hypothesis that all three
slopes are the same, or equal. The significance level of
5 percent indicates that if all three slopes are the same,
then there is a 5 percent chance that the curves would be
as different as they are. Therefore, the hypothesis is
discarded; and the slopes are considered different.

5. t-Test of Correlation Coefficients

In order to assess which of the three products agreed most closely with the data base, a t-test of correlation coefficients was performed for the forest and agriculture/ rangeland products. The test was not performed on the water class because the three curves (products) were determined to be the same in the analysis of covariance. The urban class was not tested because none of the three curves yielded a statistically significant fit.

The basis for this test is the correlation coefficient, which when squared is an estimate of the percent reduction of the error due to the regression. The test statistic was:

$$t = \left(|R_{y4}| - |R_{y2}| \right) \left(\sqrt{\frac{(n-3)(1 + |R_{12}|)}{2D}} \right)$$

where

$$D = 1 + 2 \left(|R_{12}| \cdot |R_{y1}| \cdot |R_{y2}| \right) - \left(R_{12}^2 + R_{y1}^2 + R_{y2}^2 \right)$$

For the forest class, the two best products were the computer and JSC color composite. The t-test indicated that the computer was significantly better than the JSC color composite (significant at the 0.1-percent level).

For the agriculture/rangeland class the two best products were again the computer and JSC color composite. The t-test indicated that the computer was better than the JSC color composite but at a significance level of only 20 percent. However, for further analysis the computer was selected as the better.

6. Analysis of Covariance to Determine One-to-One Relationship

As a result of the regression analysis (section 3 of appendix B), an equation was produced for every class for each product. The equations can be found in table C-1 of appendix C.

The purpose of the equations is to increase the accuracy of the initial occurrence estimates. Covariance was performed in order to determine if there is actually no significant difference between the initial occurrence estimates and the estimates obtained by the equation.

The test was performed on the computer product for the forest and agriculture/rangeland classes, as well as the combined (three-product) regression for the water class. The results indicated that each of the three corrected estimators (equations) was significantly different from the original estimates.

APPENDIX C

STATISTICAL ANALYSIS OF LAND-USE DATA BASED
ON A STRATIFIED STATISTICAL SAMPLING SCHEME

STATISTICAL ANALYSIS OF LAND-USE DATA BASED ON
A STRATIFIED STATISTICAL SAMPLING SCHEME

The analyses presented in this appendix utilize the original 168 quadrats described at the beginning of appendix A. The same methods of statistical analyses as those demonstrated in appendix B are employed. However, a stratified statistical sampling of the original quadrat data is introduced.

1. Agreement Between Interpreters on Data Base Compilation

The determination of occurrence of an item (class) in an area is a binomial distribution; it occurs or it does not occur. The standard deviation of a binomial distribution is theoretically described as $\sigma = \sqrt{pq/n}$ where p is the probability of finding the class; $q = 1 - p$, or the probability of not finding the class; and n is the number of trials, or attempts to find a class.

An 85-percent confidence on the agreement between two interpreters on the occurrence of a class within an area dictates the following thresholding values:

TABLE C-1.— THRESHOLDS FOR DISCARDING QUADRATS FROM ANALYSIS

Average of the Two Interpreters for a Quadrat	Difference Between Cannot Exceed, Percent
0-19	6
20-39	9
40-59	10
60-79	9
80-100	6

This technique eliminated very few of the quadrats and kept most of the quadrats eliminated in the lower percent occurrence range by the equation used earlier.

2. Stratified Random Sample for Regression

For purposes of sampling for regression analysis, an
equally likely distribution of points over the desired range
of the regression should be establisned. Therefore, five
equal intervals over a range of 0 percent through 100 percent
(symmetrical chart 10 percent, 30 percent, 50 percent,
70 percent, and 90 percent) were used. The total number of
samples (quadrats) should be 30 for the purpose of getting
out of the range of small sample statistics.

Each quadrat still available for sampling was assigned
a number. A table of random numbers was then used to select
a quadrat to be used in the regression. Six quadrats were
selected per interval. (See table C-2.) The water category
could not be sampled in this way because it only had six
quadrats total. Therefore, this type analysis was not per-
formed on this category.

The urban category class was not evaluated in this manner
because there were no quadrats available with less than
80 percent or more than 19 percent urban.

TABLE C-2.— DISTRIBUTION OF QUADRATS SELECTED IN
STRATIFIED SAMPLING SCHEME

Occurrence of a Feature in a Quadrat, Percent	Forest	Agriculture/Range
80-100	6	6
60-79	6	6
40-59	6	6
20-39	6	6
0-19	6	6
Total	30	30

3. Statistical Analysis

The forest and agriculture/range classes were then
evaluated with respect to the three products in the same
way as mentioned earlier with the results contained in
table C-3 and figure C-1.

C-2

TABLE C-3.— EQUATION, STANDARD ERROR, AND SIGNIFICANCE FOR FOREST AND AGRICULTURE/RANGE
BLACK-AND-WHITE IMAGERY, JSC COLOR COMPOSITE, AND COMPUTER

	Black-and-White Imagery			JSC Color Composite			Computer		
	Equation	Standard Error	Significance	Equation	Standard Error	Significance	Equation	Standard Error	Significance
Forest	$8.7+0.83*E$	15.0	>1% $R=0.888$ $R^2=0.789$	$5.2+0.76*E$	15.4	>1% $R=0.878$ $R^2=0.771$	$8.3+0.95*E$	7.2	>1% $R=0.975$ $R^2=0.951$
Agriculture/ Range	$17.5+0.67*E$	17.4	>1% $R=0.864$ $R^2=0.746$	$17.0+0.68*E$	15.4	>1% $R=0.880$ $R^2=0.774$	$-3.3+0.97*E$	16.7	>1% $R=0.859$ $R^2=0.738$

NOTE: E = estimate obtained from product; R = correlation coefficient; R^2 = percent reduction error due to regression.
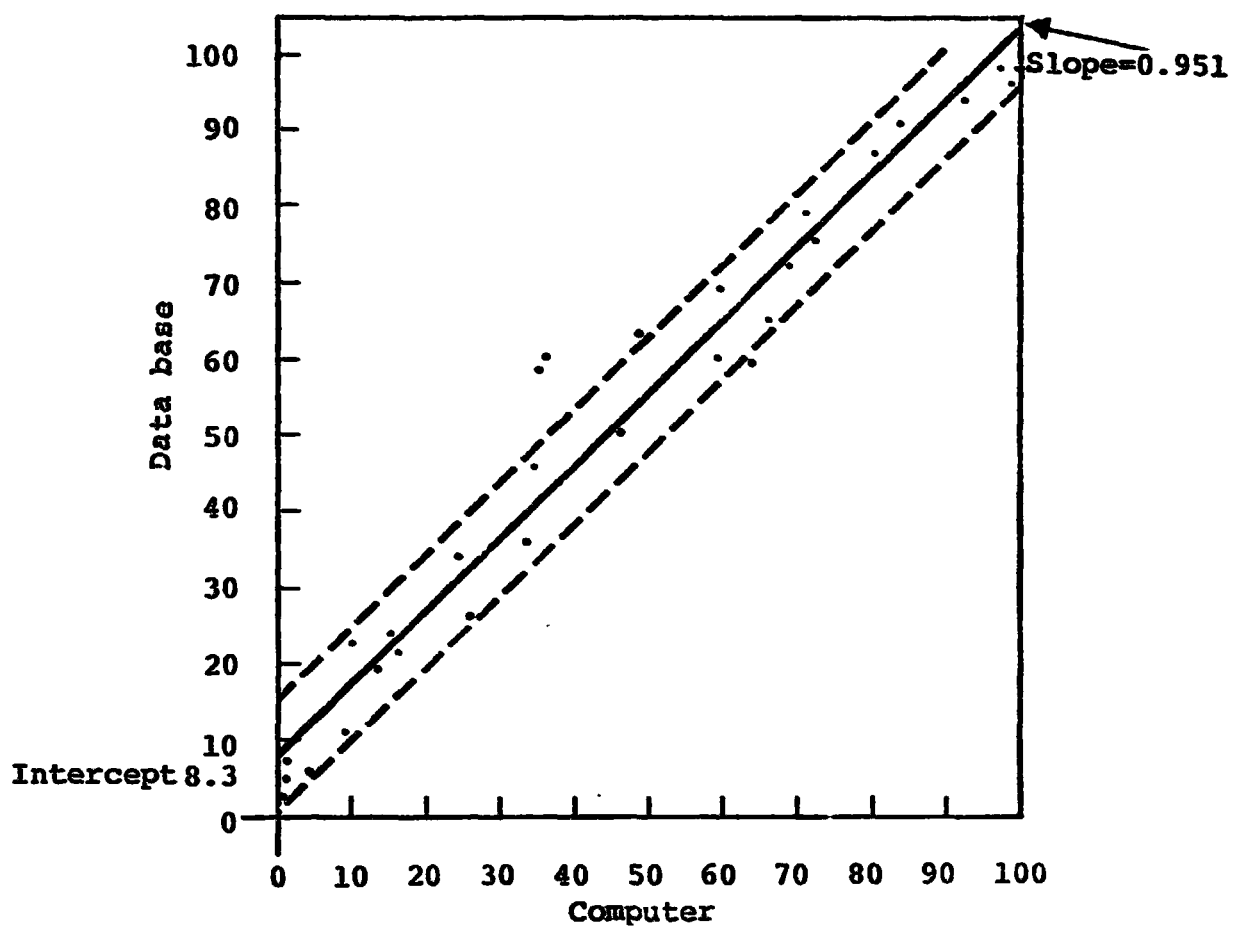
Figure C-1.— Least-squares regression of forest occurrence
for computer product on forest occurrence for data base,
utilizing the sampled quadrats.

4. Analysis of Covariance

No analysis of covariance was performed for the three products for the forestry category since the slope for the computer techniques was substantially different from the other two products.

The statistical difference between the regression curves for the agriculture/range products indicated a significance of >5 percent for the slopes. Therefore, a difference is assumed to exist between the three products.

5. Correlation Coefficients

The two best curve fits or two best products for the forest category were the computer and the black-and-white imagery. The computer was the better product by a significance of better than 1 percent.

The two best products for the agriculture/range category were the computer and the JSC color composite imagery. The t-test indicated that there was no difference between the two products (the difference was insignificant) with regard to the agreement and precision of the regression equations. Since the computer product equation approached a one-to-one relationship with respect to the data base, the computer product was a one-to-one relationship.

6. Analysis of Covariance for a One-to-One Relationship

The computer product for the forest category differed from a direct relationship with the data base with a significance that was >1 percent. Therefore, the equation is still required.

The computer product for the agriculture/rangeland category differed from a direct relationship with the data base with a significance that was >10 percent. Less than 10 percent is a lack-of-confidence region of the significance data in determining if a difference exists between the regression model for the product and a true one-to-one relationship.

The recommended procedure for either strengthening the confidence in the model or completely discarding it is to increase the six sample sites and do the analysis again. Another alternative is to use both techniques on other agriculture/range areas in a sequential test to determine which is the better.

APPENDIX D

GLOSSARY AND ACRONYMS

band

> A group of wavelengths of light producing one color or convenient group of wavelengths, such as near-infrared.

channel

> The same as "band" when used in computer work.

clustering

> Mathematical procedure for organizing multispectral data into spectrally homogeneous groups. Clusters require identification and interpretation in a post-processing analysis. ISOCLS is a spectral clustering program.

color composite

> Color composite of three channels of ERTS-1 multi-spectral scanner digital data. The composites are third- or fourth-generation images, compared to first-generation composites produced from computer-compatible tapes using a film recorder.

computer-compatible tapes

> Tapes containing digital ERTS-1 data. These tapes are standard 19-cm (7-1/2-in.) wide magnetic tapes in 9-track or 7-track format. Four tapes are required for the four-band multispectral digital data corresponding to one ERTS-1 scene.

DAS

> Data analysis station, a computer system of tape drives and computer, a display and control console, and film recorder. The DAS is used to reformat, analyze, and review remotely sensed digital data tapes.

DLMIN

The minimum distance threshold for combining clusters.

E

Estimate obtained from product. This factor appears
in the equations of the black and white imagery, JSC
color composite, and computer.

EMBEDT

Univac 1108 program designed primarily to convert the
ERTS system-corrected tape produced by Goddard Space
Flight Center to multispectral data system edit format.

ERTS-1 scene

Collection of the image data of one nominal framing
area (185 km^2) of the Earth's surface. The scene
includes all data from each spectral band of each
sensor.

gray scale

A scale of gray tones between white and black with an
arbitrary number of segments. The ERTS-1 images have
a 15-step gray scale exposed on every frame of imagery.
The scale gives the relationship between gray level on
the image and the electron beam density used to expose
the original image.

ha

Hectare, a metric unit of area equal to 10,000 m^2 or
2.47 acres.

ISOCLS

Iterative Self-Organizing Clustering System, a computer
program developed at JSC using a clustering algorithm
to group homogeneous spectral data. Controlling inputs
allow investigators to control the size and number of

clusters. Because the system produces a classification-
type clustering map in which clusters require post-
processing identification and interpretation, the system
is frequently called a nonsupervised classification
system.

LARSYS

The set of classification programs for aircraft data
handling and analysis developed at the Laboratory for
the Applications of Remote Sensing, Purdue University.

maximum likelihood ratio

Maximum likelihood ratio in remote sensing is a prob-
ability decision rule for classifying a target from
multispectral data. Two types of errors are feasible:
failure to classify the target correctly and mis-
classification of background as the target. In its
simplest form, the likelihood ratio is P_t/P_b. This
expression compares the probability (P) of an unknown
spectral measurement being classified as target (t) to
the probability of an unknown spectral measurement
being classified as background (b). When $P_t/P_b \geq 1$,
the formula decides t; and when P $P_t/P_b < 1$, it decides
b. Probability density functions are computed from
spectral samples, often called training samples. As
the number of training samples increases, the mathe-
matical computations of the maximum likelihood ratio
increase in complexity. As a result, digital computer
analysis is required. The analysis is called automatic
data processing of multispectral remotely sensed data
or automatic spectral pattern recognition of multi-
spectral remotely sensed data.

MSS

Multispectral scanner system, sometimes called the
multispectral scanner. The MSS usually refers to
the ERTS-1 operational scanning system.

multiband

A study using more than one band.

multispectral scanner spectral bands

The division of the visible and near infrared portions
of the electromagnetic spectrum into discrete segments.

MSS channel	ERTS-1 band	Wavelength, nm	Color
1	4	500-600	green
2	5	600-700	red
3	6	700-800	reflective
4	7	800-1,100	infrared

nmi

Nautical mile, equalling 1/60th of a degree at the
Earth's equator, or about 6,076 ft.

nonsupervised classification

A procedure grouping spectral data into homogeneous
clusters. Identification and interpretation are done
in a postprocessing analysis.

pixel

Picture resolution element, or one instantaneous field
of view recorded by the multispectral scanning system.
An ERTS-1 pixel is about 0.44 hectare (1.09 acres).
One ERTS-1 frame contains abcut 7.36×10^6 pixels, each
described by four radiance values.

R

    Correlation coefficient. This factor appears in the
significance of black and white imagery, JSC color
composite, and computer.

R^2

    Percent reduction error due to regression. This factor
appears in the significance of black and white imagery,
JSC color composite, and computer.

radiance

    Measure of the radiant energy emitted by a radiator
in a given direction.

reflectance

    Ratio of the radiance of the energy reflected from a
body to that incident upon it. Reflectance is usually
measured in percent.

scene-corrected data

    System-corrected data processed to produce precision
located and corrected imagery on 24-cm (9-1/2-in.)
film.

signature

    A set of spectral, tonal, or spatial characteristics
of a classification serving to identify a feature by
remote sensing.

spectral response

    Spectral radiance of an object sensed at the satellite
and recorded by the multispectral scanner.

STDMAX

    The value of the standard deviation before a class is
split into two groups.

supervised classification
> Classification procedure in which data of known
> classes are used to establish the decision logic
> from which unknown data are assigned to the classes.
> The automatic data processing supervised classifica-
> tion procedure used at JSC during the ERTS-1 project
> used a Gaussian maximum likelihood decision rule.

threshold
> The boundary in spectral space beyond which a data
> point, pixel, has such a low probability of inclusion
> in a given class that the pixel is excluded from that
> class.

training field
> The spatial sample of digital data of a known ground
> feature selected by the investigator. From the sample
> the spectral characteristics are computed for super-
> vised multispectral classification of remotely sensed
> data. The statistics associated with training fields
> form the input to the maximum likelihood ratio compu-
> tations and train the computer to discriminate between
> samples.